รายการอ้างอิง

ภาษาไทย

เกรียงศักดิ์ ศิวะสนธิวัฒน์. การศึกษาการรู้จำเสียงพูดโดยตรงจากเครื่องเข้ารหัส G.729. ปริญญานิพนธ์ สาขาวิศวกรรมไฟฟ้า จุฬาลงกร์มหาวิทยาลัย, 2539.

พีรพล ทินกรศรีสุภาพ. การศึกษาการดึงจุดเด่นเชิงความถี่ของเสียงพูดจากเครื่องเข้ารหัส G.729. ปริญญานิพนธ์ สาขาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย, 2540.

วิศรุต อาขุบุตร. ระบบรู้จำคำไทยหลายพยางค์แบบไม่ขึ้นกับผู้พูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟ. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย, 2539.

เสาวลักษณ์ อารีย์พงศา. การรู้จำเสียงพูดตัวเลขเป็นภาษาไทยแบบไม่ขึ้นกับผู้พูดโดยวิธีฮิดเดนมาร์คอฟโมเดล และเวกเตอร์ควอนไตซ์เซชั่น. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย, 2538.

ภาษาอังกฤษ

Benyassine, A., Shlomot, E., Su H.Y., ITU-T Recommendation G.729 Annex B : A Silence Compression Scheme for Use with G.729 Optimized for V.70 Digital Simultaneous Voice and Data Applications, IEEE Communication Magazine (September 1997): 64-73.

Campbell, J.,Speaker Recognition : A tutorial, Proceedings of the IEEE, (September 1997) : 1436-1461.

Coetzee, H.J., Barnwell T.P., An LSP based speech quality measure, ICASSP-89 , (1989) : 596-599.

Deller, J. , Proakis, J., Hansen J.Discrete-Time Processing of Speech Signals.(331-332) : Macmillan 1993.

Evangelos, S.D., Nikos, D.F.,George, K.K. . Fast Endpoint Detection Algorithm for Isolated Word Recognition in Office Environment. IEEE Transaction on Speech and Audio processing (1991) : 733-736.

Furui, S., Sondhi, M. M.. Advances in Speech Signal Processing . Tokyo : Tokai University Press, 1985.

GerHard, S.. The Road to G.729 : ITU 8-kb/s Speech Coding Algorithm with Wireline Quality, IEEE Communications Magazine (September 1997) : 48-54.

Hamada, M., Takizawa, Y.,Norimatsu, T., A noise robust speech recognition system , Proceeding of the ICSLP-90 (1990) :893-896.

ITU-T RECOMMENDATION G.729. Coding of Speech at 8 kbit/s using Conjugate-Structure Algebraic-Code-Excited Linear Prediction(CS-ACELP). 1996.

ITU-T RECOMMENDATION G.729 (Annex B) Coding of Speech at 8 kbit/s using Conjugate-Structure Algebraic-Code-Excited Linear Prediction(CS-ACELP) Annex B : A silence compression scheme for G.729 optimized for terminals conforming to Recommendation V.70 September 1996.

Rabiner, L., Juang, B.H. Fundamental of Speech Recognition.(267-270) : Prentice-Hall,1993.

Rabiner, L. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proceeding of the IEEE (February 1989) : 257-286.

Salami, R., Laflamme, C.,Bessette, B., Adoul, J.P. ITU-T B.729 Annex A : Reduced Complexity 8 kb/s CS-ACELP Codec for Digital Simultaneous Voice and Data, IEEE Communication Magazine (September 1997) : 56-63.

Schalkoff, R. Pattern Recognition : Statistical, Structural and Neural approaches. John Wiley & Sons (1992).

Smith, A.C., Schalkwyk, V. Line-Spectrum Pairs – A Review, COMSIG 88 (1988) : 7-11.

Ying, G.S.,Mitchell, C.D.,Jamieson, L.H. Endpoint Detection of isolated utterances based on a modified teager energy measurement, IEEE Transaction on Speech and Audio processing (1993) : 732-735.

ภาคผนวก

ภาคผนวก ก

รายการชุดคำศัพท์ภาษาไทย

ในภาคผนวกนี้จะกล่าวถึงรายละเอียดของชุดคำศัพท์ภาษาไทยจำนวน 30 คนแบ่งเป็น 2 ชุดได้แก่ ชุดคำศัพท์ตัวเลขศูนย์ถึงเก้า จำนวน 10 คำ และชุดคำศัพท์พยางค์เดียว จำนวน 20 คำ

ตารางที่ ก.1 รายละเอียดชุดคำศัพท์ตัวเลขศูนย์ถึงเก้า

คำศัพท์ตัวเลขศูนย์ถึงเก้า	สัญลักษณ์แทนการออกเสียง
หนึ่ง	/'hnvng1/
สอง	/'s@@ng4/
สาม	/'saam4'/
สี่	/'sii1/
ห้า	/'haa2/
หก	/'hok1/
เจ็ด	/'cet1/
แปด	/'pxxt1/
เก้า	/'kaw2/
ศูนย์	/'suun4/

ตารางที่ ก.2 รายละเอียดชุดคำศัพท์พยางค์เดียว

คำศัพท์พยางค์เดียว	สัญลักษณ์แทนการออกเสียง
เดิน	/'dqqn0/
วิ่ง	/'wing2/
นอน	/'n@@n0/
ตา	/'taa0/
ปาก	/'paak1/
หู	/'huu4/
มือ	/'mvv0/
เทียน	/'thiian0/
กิน	/'kin0/
นก	/'nok3/
เป็ด	/'pet1/
ไก่	/'kaj1/
กล้วย	/'kluuaj2/
ส้ม	/'som2/
โต๊ะ	/'to3/
เตียง	/'tiiang0/
นั่ง	/'nang2/
แก้ว	/'kxxw2/
น้ำ	/'nam3/
เสือ	/'svva4/

ภาคผนวก ข

ตารางผลการรู้จำคำศัพท์ตัวเลขและคำศัพท์พยางค์เดียว

ในภาคผนวกนี้จะกล่าวถึงรายละเอียดของผลการรู้จำชุดคำศัพท์ตัวเลขศูนย์ถึงเก้า จำนวน 10 คำ และชุดคำศัพท์พยางค์เดียว จำนวน 20 คำ

ตารางที่ ข.1 รายละเอียดผลการรู้จำชุดเสียงพูดทดสอบคำศัพท์ตัวเลขศูนย์ถึงเก้าวิธีที่ 1 ชุดที่5

	ผลเสียงพูด	หนึ่ง	สอง	สาม	สี่	ห้า	หก	เจ็ด	แปด	เก้า	ศูนย์	ร้อยละของผลการรู้จำ
เสียงพูดทดสอบ												
หนึ่ง		20	0	0	0	0	0	0	0	0	0	100
สอง		0	20	0	0	0	0	0	0	0	0	100
สาม		0	0	20	0	0	0	0	0	0	0	100
สี่		0	0	0	20	0	0	0	0	0	0	100
ห้า		0	0	3	0	17	0	0	0	0	0	85
หก		0	1	0	0	0	18	0	0	0	1	90
เจ็ด		0	0	0	3	0	0	17	0	0	0	85
แปด		0	0	0	0	3	0	0	17	0	0	85
เก้า		0	0	2	0	0	0	0	0	18	0	90
ศูนย์		0	0	1	0	0	0	0	0	0	19	95

ผลการรู้จำถูกต้องร้อยละ 93.00

ตารางที่ ข.2 รายละเอียดผลการรู้จำชุดเสียงพูดทดสอบคำศัพท์ตัวเลขศูนย์ถึงเก้าวิธีที่ 2 ชุดที่5

ผลเสียงพูด / เสียงพูดทดสอบ	หนึ่ง	สอง	สาม	สี่	ห้า	หก	เจ็ด	แปด	เก้า	ศูนย์	ร้อยละของผลการรู้จำ
หนึ่ง	20	0	0	0	0	0	0	0	0	0	100
สอง	0	20	0	0	0	0	0	0	0	0	100
สาม	0	0	20	0	0	0	0	0	0	0	100
สี่	0	0	1	19	0	0	0	0	0	0	95
ห้า	0	0	2	0	16	0	0	1	0	1	80
หก	0	0	0	0	0	18	0	0	0	2	90
เจ็ด	0	0	0	3	0	0	16	0	0	1	80
แปด	0	0	1	1	2	0	0	15	1	0	75
เก้า	0	1	1	0	0	0	0	0	17	1	85
ศูนย์	0	1	0	0	0	0	1	0	0	18	90

ผลการรู้จำถูกต้องร้อยละ 89.50

ตารางที่ ข.3 รายละเอียดผลการรู้จำชุดเสียงพูดทดสอบคำศัพท์ตัวเลขศูนย์ถึงเก้าวิธีที่ 3 ชุดที่5

ผลเสียงพูด / เสียงพูดทดสอบ	หนึ่ง	สอง	สาม	สี่	ห้า	หก	เจ็ด	แปด	เก้า	ศูนย์	ร้อยละของผลการรู้จำ
หนึ่ง	18	0	0	1	0	0	1	0	0	0	90
สอง	0	19	0	0	0	0	0	0	0	1	95
สาม	0	0	20	0	0	0	0	0	0	0	100
สี่	0	0	0	20	0	0	0	0	0	0	100
ห้า	0	0	0	0	18	0	0	2	0	0	90
หก	0	1	0	0	0	18	0	0	0	1	90
เจ็ด	0	0	0	2	0	0	18	0	0	0	90
แปด	0	0	0	0	1	0	1	18	0	0	90
เก้า	0	0	1	0	1	0	0	0	18	0	90
ศูนย์	0	0	0	0	0	0	1	0	0	19	95

ผลการรู้จำถูกต้องร้อยละ 93.00

ตารางที่ ข.4 รายละเอียดผลการรู้จำชุดเสียงพูดทดสอบคำศัพท์ตัวเลขศูนย์ถึงเก้าวิธีที่ 4 ชุดที่ 4

	ผลเสียงพูด	หนึ่ง	สอง	สาม	สี่	ห้า	หก	เจ็ด	แปด	เก้า	ศูนย์	ร้อยละของผลการรู้จำ
เสียงพูดทดสอบ												
หนึ่ง		20	0	0	1	0	0	1	0	0	0	100
สอง		0	20	0	0	0	0	0	0	0	0	100
สาม		0	0	20	0	0	0	0	0	0	0	100
สี่		0	2	0	18	0	0	0	0	0	0	90
ห้า		0	0	1	0	19	0	0	0	0	0	95
หก		0	2	0	0	0	13	0	0	0	5	65
เจ็ด		1	0	0	4	0	0	15	0	0	0	75
แปด		1	1	0	0	1	0	0	17	0	0	85
เก้า		0	0	1	0	0	0	0	1	18	0	90
ศูนย์		0	1	0	0	0	0	1	0	0	18	90

ผลการรู้จำถูกต้องร้อยละ 89.00

ตารางที่ ข.5 รายละเอียดผลการรู้จำชุดเสียงพูดทดสอบคำศัพท์พยางค์เดียว วิธีที่ 1 ชุดที่ 5

ผลเสียงพูด / เสียงพูดทดสอบ	เดิน	วิ่ง	นอน	ตา	ปาก	หู	มือ	เทียน	กิน	นก	เป็ด	ไก่	กล้วย	ส้ม	โต๊ะ	เตียง	นั่ง	แก้ว	น้ำ	เสือ	ร้อยละของผลการรู้จำ
เดิน	19	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	95
วิ่ง	0	19	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	95
นอน	0	0	18	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	90
ตา	0	0	1	13	2	0	0	0	0	0	0	2	0	0	0	0	1	0	1	0	65
ปาก	0	0	1	4	13	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	65
หู	0	0	0	0	0	18	0	0	0	0	0	0	1	0	0	0	0	1	0	0	90
มือ	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0	1	95
เทียน	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	100
กิน	0	0	0	0	0	0	0	2	18	0	0	0	0	0	0	0	0	0	0	0	90
นก	0	0	2	0	0	1	0	0	0	17	0	0	0	0	0	0	0	2	0	0	85
เป็ด	0	0	0	0	0	1	0	0	0	0	13	1	0	0	0	3	0	0	0	0	65
ไก่	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	100
กล้วย	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	100
ส้ม	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	100
โต๊ะ	0	0	0	0	0	0	0	0	0	0	0	0	1	3	16	0	0	0	0	0	80
เตียง	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	18	0	0	0	0	90
นั่ง	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	16	0	3	0	80
แก้ว	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	100
น้ำ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	18	1	90
เสือ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	19	95

ตารางที่ ข.6 รายละเอียดผลการรู้จำชุดเสียงพูดทดสอบคำศัพท์พยางค์เดียว วิธีที่ 2 ชุดที่ 5

ผลเสียงพูด / เสียงพูดทดสอบ	เดิน	วิ่ง	นอน	ตา	ปาก	หู	มือ	เทียน	กิน	นก	เป็ด	ไก่	กล้วย	ส้ม	โต๊ะ	เตียง	นั่ง	แก้ว	น้ำ	เสือ	ร้อยละของผลการรู้จำ
เดิน	17	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	85
วิ่ง	0	18	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	90
นอน	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100
ตา	0	0	1	11	5	0	0	0	0	0	0	1	0	0	0	0	2	0	0	0	55
ปาก	0	0	0	6	13	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	65
หู	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100
มือ	0	0	0	0	0	0	18	1	0	0	0	0	0	0	0	0	1	0	0	0	90
เทียน	0	1	0	0	0	0	0	16	3	0	0	0	0	0	0	0	0	0	0	0	80
กิน	0	0	0	0	0	0	0	4	16	0	0	0	0	0	0	0	0	0	0	0	80
นก	0	0	4	0	0	0	0	0	0	13	0	0	1	2	0	0	0	0	0	0	65
เป็ด	0	0	1	0	0	0	0	4	0	1	12	0	0	0	0	1	1	0	0	0	60
ไก่	0	0	1	0	0	0	1	0	0	0	0	18	0	0	0	0	0	0	0	0	90
กล้วย	0	0	1	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0	0	95
ส้ม	0	0	2	0	0	0	0	0	0	0	0	0	0	18	0	0	0	0	0	0	90
โต๊ะ	0	0	1	0	0	0	0	0	0	0	0	0	0	1	18	0	0	0	0	0	90
เตียง	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	16	0	0	1	0	80
นั่ง	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	4	0	80
แก้ว	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	100
น้ำ	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2	1	15	1	75
เสือ	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	19	95

ตารางที่ ข.7 รายละเอียดผลการรู้จำชุดเสียงพูดทดสอบคำศัพท์พยางค์เดียว วิธีที่ 3 ชุดที่ 5

ผลเสียงพูด เสียงพูดทดสอบ	เดิน	วิ่ง	นอน	ตา	ปาก	หู	มือ	เทียน	กิน	นก	เป็ด	ไก่	กล้วย	ส้ม	โต๊ะ	เตียง	นั่ง	แก้ว	น้ำ	เสือ	ร้อยละของผลการรู้จำ
เดิน	18	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	90
วิ่ง	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100
นอน	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	95
ตา	1	0	0	14	3	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	70
ปาก	0	0	0	5	11	0	0	0	0	0	0	2	0	1	0	0	0	1	0	0	55
หู	0	0	0	0	0	19	0	0	0	0	0	0	1	0	0	0	0	0	0	0	95
มือ	2	1	1	0	0	0	15	1	0	0	0	0	0	0	0	0	0	0	0	0	75
เทียน	0	0	0	0	0	0	0	17	1	0	0	0	0	0	0	2	0	0	0	0	85
กิน	0	0	0	0	0	0	0	2	18	0	0	0	0	0	0	0	0	0	0	0	90
นก	1	0	2	0	0	0	1	0	0	15	0	0	1	0	0	0	0	0	0	0	75
เป็ด	0	0	0	0	0	0	0	2	0	0	8	1	0	0	0	4	0	5	0	0	40
ไก่	0	0	0	0	0	0	0	0	0	0	0	19	1	0	0	0	0	0	0	0	95
กล้วย	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	100
ส้ม	0	0	1	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0	95
โต๊ะ	0	0	0	0	0	0	0	0	0	0	0	0	0	4	16	0	0	0	0	0	80
เตียง	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	17	0	0	0	0	85
นั่ง	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	6	1	65
แก้ว	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	100
น้ำ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	15	2	75
เสือ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	19	95

ตารางที่ ข.8 รายละเอียดผลการรู้จำชุดเสียงพูดทดสอบคำศัพท์พยางค์เดียว วิธีที่ 4 ชุดที่ 4

ผลเสียงพูด เสียงพูดทดสอบ	เดิน	วิ่ง	นอน	ตา	ปาก	หู	มือ	เทียน	กิน	นก	เป็ด	ไก่	กล้วย	ส้ม	โต๊ะ	เตียง	นั่ง	แก้ว	น้ำ	เสือ	ร้อยละของผลการรู้จำ
เดิน	15	0	1	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	1	0	75
วิ่ง	0	18	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	90
นอน	2	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	90
ตา	0	0	0	15	3	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	75
ปาก	0	0	0	9	9	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	45
หู	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100
มือ	3	0	0	0	0	0	15	1	0	0	0	0	0	0	0	1	0	0	0	0	75
เทียน	0	0	0	0	0	0	0	14	1	0	0	0	0	0	0	5	0	0	0	0	70
กิน	0	0	0	0	0	0	0	1	18	0	0	0	0	0	0	1	0	0	0	0	90
นก	1	0	1	0	0	0	0	0	0	13	0	0	3	1	0	1	0	0	0	0	65
เป็ด	1	2	0	0	0	0	0	2	0	0	8	1	1	0	0	3	0	2	0	0	40
ไก่	0	0	0	1	0	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0	95
กล้วย	0	0	1	0	0	1	0	0	0	0	0	0	17	0	0	0	1	0	0	0	85
ส้ม	0	0	0	0	0	0	1	0	0	0	0	0	1	17	0	0	1	0	0	0	85
โต๊ะ	0	0	0	0	0	0	0	0	0	1	0	0	3	5	11	0	0	0	0	0	55
เตียง	1	0	0	0	0	0	0	3	0	0	0	0	0	0	0	15	0	0	0	1	75
นั่ง	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	9	3	6	0	45
แก้ว	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	100
น้ำ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	5	2	12	0	60
เสือ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	100

ภาคผนวก ค

บทความที่ได้รับการเผยแพร่

ในภาคผนวกนี้ประกอบไปด้วยบทความที่ได้รับการเผยแพร่ ซึ่งได้แก่ การศึกษาการดึงจุดเด่นเชิงความถี่ของเสียงพูดโดยตรงจากการเข้ารหัส G.729 (Study of Spectral Feature Extraction Directly from G.729 Coded Speech

การศึกษาการดึงจุดเด่นเชิงความถี่ของเสียงพูดโดยตรงจากการเข้ารหัส G.729

Study of Spectral Feature Extraction Directly from G.729 Coded Speech

พีรพล ทินกรศรีสุภาพ , สิริ วงศ์วรชาติกาล และ สุวิทย์ นาคพีระยุทธ
ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ถนนพญาไท ปทุมวัน กรุงเทพฯ 10330
โทร. (02) 218-6512 E-mail: suvit@ee.eng.chula.ac.th

บทคัดย่อ

มาตรฐาน ITU-T G.729 เป็นมาตรฐานในการบีบอัดเสียงพูด ซึ่งสามารถนำใช้งานได้อย่างกว้างขวาง ดังนั้นถ้าเราสามารถดึงจุดเด่นของเสียงที่จำเป็นในการรู้จำออกมาได้โดยตรงจากรหัสเสียงที่ถูกบีบอัดแล้ว จะสามารถสร้างระบบรู้จำเสียงอย่างง่ายจากรหัสเสียง G.729 โดยตรง ความถี่ LSP เป็นพารามิเตอร์ตัวหนึ่งที่ส่งมากับรหัส G.729 และเป็นตัวแทนของผลตอบเชิงความถี่ของเสียงพูดที่ได้รับการยอมรับอย่างกว้างขวางในงานรู้จำเสียง จากการศึกษาลักษณะสมบัติของความถี่ LSP พบว่าช่วงที่ความถี่ LSP รวมกลุ่มกันอยู่หนาแน่นจะเป็นช่วงความถี่ฟอร์แมนท์ของ vocal tract ด้วย บทความนี้นำเสนอการทดสอบการหาความถี่ฟอร์แมนท์ด้วยการจัดกลุ่มความถี่ LSP แบบต่างๆเปรียบเทียบกับสเปคโตรแกรมและผลตอบเชิงความถี่ที่แท้จริงของ LSP และนำผลที่สังเกตได้ไปสร้างเป็นระเบียบวิธีอย่างง่ายซึ่งลดความซับซ้อนในการคำนวณลงแต่ยังคงความแม่นยำในการประมาณค่าความถี่ฟอร์แมนท์ โดยระเบียบวิธีที่ได้นี้จะสามารถใช้ในการศึกษาการรู้จำเสียงโดยตรงจากรหัสเสียง G.729 ได้ต่อไป

Abstract

The ITU-T Recommendation G.729 is a versatile and well accepted speech compression standard. If the speech feature can be extracted directly from the code easily, a simple speech recognition system can work directly on the G.729 codes LSP is one of the parameters obtained from G.729 codes which contains the speech spectral information neccessary for speech recognition. The study of LSP characteristics revealed that the vocal tract formant frequencies can be found at each LSP clustering frequencies. The approximation of formant frequencies by several clustering algorithms of the LSP was tested in this article. The test results were then compared to the spectrogram and the actual frequency responses of LSP. Then the heuristic algorithm, based on the clustering method, with low complexities was suggested. The use of G.729 speech coder as a preprocessor for speech recognition system can be studied further based on this algorithm.

1. บทนำ

มาตรฐาน ITU-T G.729 ใช้สำหรับบีบอัดเสียงพูดที่สุ่มด้วยอัตรา 8 kbit/sec ซึ่งได้ประกาศเป็นมาตรฐานสากลในปี 1996 [1] โดยคาดว่าจะมีการนำไปใช้งานอย่างกว้างขวางในระบบสื่อสารต่างๆ แสดงว่าสัญญาณเสียงพูดที่รับส่งกันจะอยู่ในรูปที่ถูกเข้ารหัสไว้แล้วเป็นส่วนใหญ่ ดังนั้นถ้าเราสามารถทำการรู้จำเสียงพูดจากรหัสเสียงโดยตรงได้โดยไม่ต้องถอดรหัสเพื่อสร้างสัญญาณเสียงเดิมกลับคืนมาก่อนแล้ว จะทำให้สามารถสร้างระบบรู้จำเสียงพูดอย่างง่ายที่มีราคาถูกได้ กระบวนการดึงจุดเด่นของเสียงพูด (Speech Features Extraction) ออกมาจากรหัสเสียง G.729 โดยตรง จึงเป็นสิ่งจำเป็นขั้นแรกต่อเป้าหมายดังกล่าว จุดเด่นที่จะดึงออกมานี้จำเป็นจะต้องใช้การคำนวณที่มีความซับซ้อนต่ำกว่าการคำนวณจากเสียงพูดโดยตรงมาก จึงจะทำให้วิธีการนี้มีข้อได้เปรียบกว่าเดิม บทความนี้จะกล่าวถึงเฉพาะการดึงจุดเด่นเชิงความถี่เท่านั้น ซึ่งเป็นส่วนที่สำคัญที่สุดในการรู้จำเสียงพูด

2. มาตรฐานการบีบอัดเสียงพูด ITU-T G.729 [1]

การเข้ารหัส G.729 ใช้หลักการของ CS-ACELP (Conjugate-Structure Algebraic-Code-Excited Linear-Predictive) การเข้ารหัสนี้ทำงานบนเฟรมเสียงความยาว 10 msec ซึ่งเท่ากับ 80 ตัวอย่าง ทุกๆเฟรมสัญญาณเสียงจะถูกวิเคราะห์เพื่อแยกพารามิเตอร์ตามแบบจำลอง CELP พารามิเตอร์เหล่านี้คือ linear predictive filter coefficients, adaptive and fixed codebook indices and gain พารามิเตอร์เหล่านี้จะถูกลงรหัสและถูกส่งไปในช่องสัญญาณ เมื่อเครื่องถอดรหัสได้รับสัญญาณก็จะใช้พารามิเตอร์เหล่านี้เพื่อสร้าง excitation และฟิลเตอร์สังเคราะห์ การสังเคราะห์เสียงพูดเริ่มจากการสร้าง excitation ขึ้นมาด้วยสัญญาณพัลส์ 4 พัลส์ที่มีขนาดและตำแหน่งตาม fixed-codebook vector quantization ร่วมกับ adaptive-codebook ซึ่งนำเอาค่า excitation ในอดีตมาใช้ซ้ำเดิมใหม่ตาม

ความเหมาะสมซึ่งขึ้นอยู่กับ pitch หรือสามารถมองในลักษณะ long-term synthesis filter ได้นั่นเอง สัญญาณ excitation จะถูกนำไปผ่าน short-term synthesis filter หลังจากนั้นเสียงที่สร้างขึ้นใหม่จะถูกปรับปรุงให้ดีขึ้นด้วย postfilter เพื่อลดผลของ artifact ที่หูมนุษย์สามารถได้ยินได้ลง ในส่วนของการทำ short-term synthesis filter นั้น มาจากการวิเคราะห์ด้วย linear predictive filter อันดับ 10 ในการวิเคราะห์สัญญาณเสียงทางด้านส่งจะใช้แบบจำลอง Linear Predictive (LP) ก่อนแล้วเปลี่ยนรูปสัมประสิทธิ์ LPC ให้เป็นค่า Line Spectrum Pair (LSP) ส่งมาแทน ซึ่งเป็นรูปแบบหนึ่งในการแสดงผลตอบเชิงความถี่ของ vocal tract ของมนุษย์ได้

3. Line Spectrum Pair

ในแบบจำลอง LP เราสามารถประมาณฟังก์ชั่นระบบของ vocal tract ได้ดังนี้

$$H(z) = \frac{G}{1 + \sum_{k=1}^{p} a_k z^{-k}} = \frac{G}{A(z)} \qquad (1)$$

โดย G คือ gain scaling factor และ A(z) คือฟิลเตอร์ผกผันของ vocal tract รูปแบบการนำเสนอฟิลเตอร์ผกผัน A(z) นั้นสามารถทำได้หลายวิธี ในการนำเสนอด้วย LSP นั้นจะทำการแปลงค่าสัมประสิทธิ์ a_k ใน (1) ให้อยู่ในโดเมนความถี่ดังนี้

$$\begin{aligned} A(z) &= \frac{1}{2}[P(z) + Q(z)] \\ P(z) &= A(z) + z^{-(p+1)}A(z^{-1}) \\ Q(z) &= A(z) - z^{-(p+1)}A(z^{-1}) \end{aligned} \qquad (2)$$

โดย p คืออันดับการทำนายของ LP ผลจากการแปลงนี้จะเกิดพหุนามสมมาตร P(z) และพหุนามอสมมาตร Q(z) ค่ารากของพหุนามทั้งสองจะมีคุณสมบัติที่น่าสนใจดังนี้ [4]

1. ศูนย์ทั้งหมดของพหุนาม LSP วางตัวอยู่บนวงกลมหนึ่งหน่วย
2. ศูนย์ของ P(z) และ Q(z) จะสลับกันไปมาอย่างมีลำดับ
3. คุณสมบัติ minimum phase ของ A(z) จะยังเป็นจริงถ้าคุณสมบัติสองข้อแรกยังคงเป็นจริงหลังการควอนไทซ์

ค่าความถี่บนวงกลมหนึ่งหน่วยของรากพหุนาม LSP นั้นจะเรียกว่าความถี่ LSP หรือ Line Spectrum Pair Frequency (LSF) กลุ่มของ LSF จะแสดงถึงรากของ A(z) หรือคือความถี่ฟอร์แมนท์ของ vocal tract ซึ่งความถี่ฟอร์แมนท์นี้จัดเป็นจุดเด่นที่สำคัญของเสียงพูดมนุษย์และมีประโยชน์มากในงานรู้จำเสียง [2]

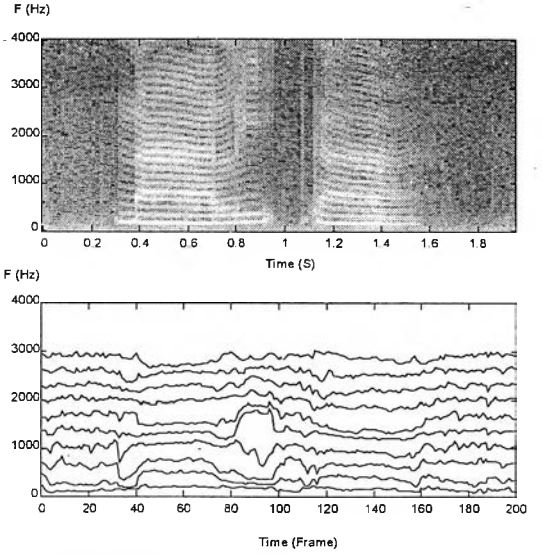
เนื่องจากความถี่ LSP เป็นพารามิเตอร์ตัวหนึ่งที่ส่งมากับรหัส G.729 ดังนั้นการนำความถี่ LSP ไปใช้ในการหาค่าความถี่ฟอร์แมนท์จึงควรจะทำได้ไม่ยากนัก

4. การทดลองและผลการทดลอง

4.1 การดำเนินการขั้นต้น

การหาความถี่ฟอร์แมนท์จากพารามิเตอร์ LSP นั้นสามารถทำได้ง่ายเพราะข้อมูลอยู่ในโดเมนความถี่โดยตรง นอกจากนี้ LSP ยังมีข้อดีอีกประการคือคงความแม่นยำได้สูงแม้จะถูกควอนไทซ์เพราะความถี่ LSP จะอยู่บนวงกลมหนึ่งหน่วยใน z-domain เสมอ

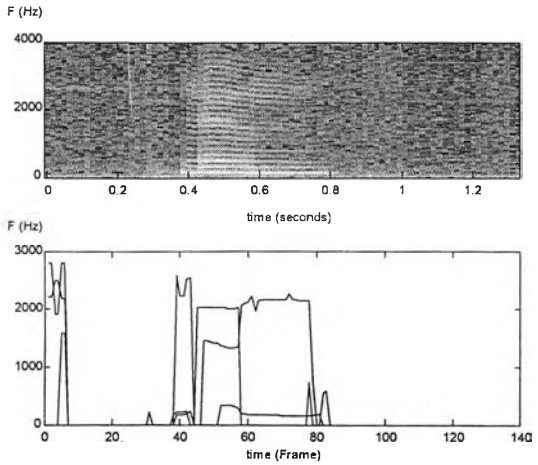
จากการวิเคราะห์ความสัมพันธ์ระหว่างความถี่ LSP กับผลตอบเชิงความถี่ในสเปคโตรแกรมแล้วจะพบว่าบริเวณที่ความถี่ LSP จับกลุ่มกันหนาแน่นจะตรงกับบริเวณที่เป็นความถี่ฟอร์แมนท์ในสเปคโตรแกรมดังรูปที่ 1 และอยู่ใกล้กับตำแหน่งขั้วของ LP เสมอ โดยขนาดของผลตอบยิ่งสูงเท่าไร LSP ก็จะอยู่ชิดกันเป็นกลุ่มมากขึ้น



รูปที่ 1 เปรียบเทียบสเปคโตรแกรมและความถี่ LSP ของคำว่า 'นาฬิกา'

ระเบียบวิธีในการจัดกลุ่ม (clustering) ที่นำมาใช้ในตอนแรกคือ วิธี Unsupervised Clustering Without Averaging (UWA) การจัดกลุ่มแบบ UWA มีความซับซ้อนต่ำ แต่ไม่รับรองการ Converge ของกลุ่มข้อมูลหรือข้อมูลที่ถูกจัดกลุ่มอาจจะไม่ครบทั้งหมด [3] แต่ความง่ายในทางปฏิบัติทำให้พิจารณาเลือกวิธีนี้ในการทดลองจัดกลุ่มความถี่ LSP โดยตรงจากพารามิเตอร์ที่แยกได้จากรหัส G.729 นั้น พบว่ามีปัญหาความไม่ต่อเนื่องเนื่องจากสัญญาณรบกวน (Background noise) ทำให้ค่าความถี่ LSP มีค่าแกว่งไปมาทั้งที่ควรมีค่าค่อนข้างคงที่ ดังนั้นจึงใช้ตัวกรองมัธยฐาน (Median Filter) กรองค่าความถี่ LSP ก่อน เพื่อช่วยทำให้

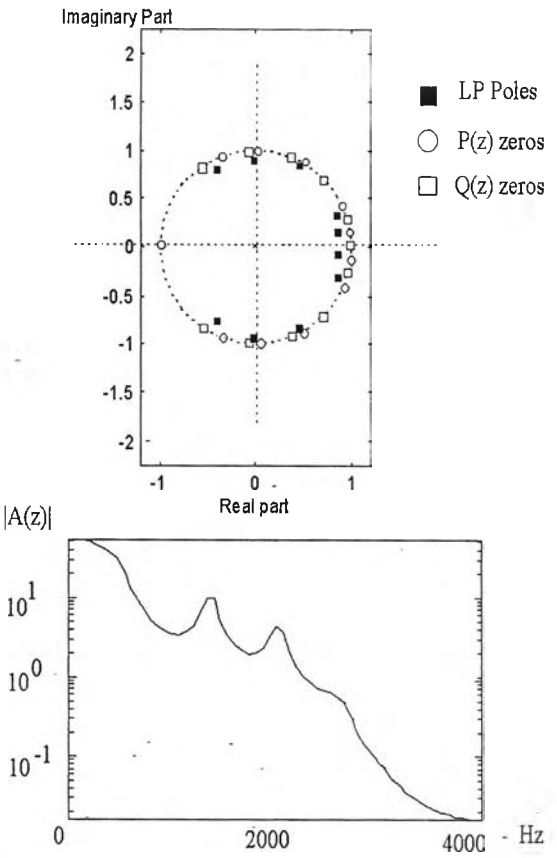
ความถี่ LSP ไม่แกว่งไปมา ผลที่ได้จากการจัดกลุ่มภายหลังการกรองมัธยฐานพบว่าได้ผลดีขึ้นกว่าเดิมดังรูปที่ 2



รูปที่ 2 เปรียบเทียบสเปคโตรแกรมและการประมาณความถี่ฟอร์แมนท์ของคำว่า "หนึ่ง"

จากการสังเกตผลตอบเชิงความถี่ที่แท้จริงของความถี่ LSP ดังรูปที่ 3 ความถี่ที่เกิดการเรโซแนนซ์ในผลตอบเชิงความถี่ก็คือความถี่ฟอร์แมนท์ของเสียงพูดนั้นจะตรงกับความถี่ที่มีขั้วของ LP อยู่ใกล้ แต่การหารากของพหุนาม LP นั้นทำได้ยากจึงไม่เหมาะที่จะทำการหาความถี่ฟอร์แมนท์จาก LP โดยตรง ส่วนการหาความถี่ ที่ความถี่ LSP รวมกลุ่มกันอย่างหนาแน่นใกล้ๆขั้วของ LP นั้นทำได้ง่ายมาก เนื่องจากความถี่ LSP อยู่ในโดเมนความถี่โดยตรงอยู่แล้ว จึงสามารถหาความถี่ฟอร์แมนท์ได้โดยง่าย นอกจากนี้ยังพบข้อสังเกตที่น่าสนใจคือ ตำแหน่งของความถี่ฟอร์แมนท์จะขึ้นอยู่กับความใกล้ชิดของความถี่ LSP เพียงสองหรือสามความถี่เท่านั้น และขึ้นกับอัตราส่วนของระยะห่างของคู่ความถี่นั้นเทียบกับคู่ความถี่อยู่ด้านข้าง

จากข้อสังเกตนี้ทำให้สามารถสร้างระเบียบวิธีการหาความถี่ฟอร์แมนท์ที่ลดความซับซ้อนลงได้อีกมาก จากการหาค่าอัตราส่วนดังกล่าวแทนการใช้การจัดกลุ่มด้วย UWA ปัญหาอีกประการหนึ่งที่พบภายหลังคือ ความไม่ต่อเนื่องของความถี่ฟอร์แมนท์ที่ได้ในทางเวลา เนื่องจากในช่วงที่ความถี่ฟอร์แมนท์มีการเปลี่ยนแปลงจะมีการเปลี่ยนกลุ่มของความถี่ LSP ทำให้การจัดกลุ่มความถี่ LSP ไม่สามารถประมาณความถี่ฟอร์แมนท์ที่เวลานั้นได้ จึงจำเป็นต้องมีการจัดกลุ่มทางเวลาเพื่อช่วยประมาณค่าความถี่บริเวณดังกล่าว และยังช่วยกำจัดผลของเสียงรบกวนออกไปได้โดยการกำจัดความถี่ฟอร์แมนท์ที่ได้ผลลัพธ์แล้วไม่เข้ากลุ่มกับเวลาข้างเคียงและ/หรือมีขนาดเล็กเกินไป



รูปที่ 3 ความถี่ LSP กับขั้วของ LP ใน z-plane พร้อมผลตอบเชิงความถี่ที่ได้

4.2 ระเบียบวิธีอย่างง่ายในการจัดกลุ่มความถี่ LSP

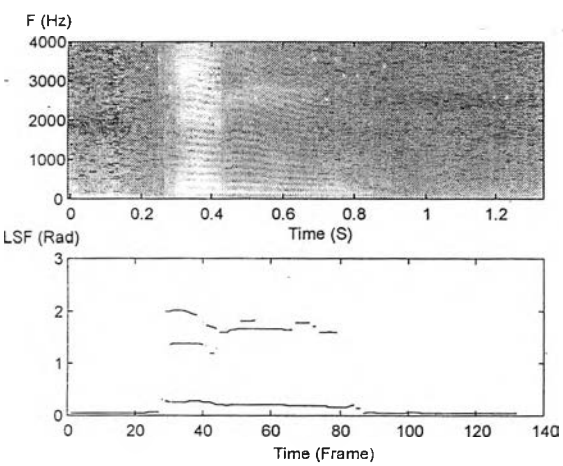
การใกล้ชิดกันของ Line Spectral Frequency (LSF) ที่ทำให้เกิดเป็นความถี่ฟอร์แมนท์นั้นสามารถประมาณได้อย่างง่ายสองแบบ คือ การใกล้ชิดกันของ LSF 2 ตัว และแบบ 3 ตัว หลักการทดสอบความใกล้กันนั้นใช้หลัก 2 ประการคือ

1. อัตราส่วนของความใกล้ชิดของคู่ LSF ที่น่าจะเกิดความถี่ฟอร์แมนท์ต่อความใกล้ชิดของคู่ LSF ที่อยู่ด้านข้าง
2. ค่าความใกล้ชิดของคู่ LSF นั้น

กรณีที่คู่ LSF 2 ตัวที่ทำให้เกิดความถี่ฟอร์แมนท์จะเกิดเมื่ออัตราส่วนของระยะห่างของ LSF คู่นั้นต่อระยะห่างของ LSF คู่ที่อยู่ติดกันมีค่าน้อยกว่า 0.6 แต่ค่าอัตราส่วนนี้จะต้องมีการปรับเปลี่ยนตามระยะห่างสัมบูรณ์ของระยะห่างของ LSF คู่นั้นด้วย โดยเมื่อระยะห่างของ LSF มีค่ามากกว่า 0.2 จากค่าที่เป็นไปได้คือ π แล้ว อัตราส่วนดังกล่าวต้องมีค่าลดลง ในทางกลับกันเมื่อระยะห่างของ LSF มีค่ามากกว่า 0.2 อัตราส่วนนี้ต้องมีค่าเพิ่มขึ้น คู่ LSF ที่สอดคล้องกับเงื่อนไขนี้จะถือเป็นตัวแทนที่น่าจะทำให้เกิดความถี่ฟอร์แมนท์ –

ส่วนการหาตัวแทนในกรณีของ LSF 3 ตัวที่ทำให้เกิดความถี่ฟอร์แมนท์นั้นก็ใช้หลักการพิจารณาแบบเดียวกัน แต่จะต้องพิจารณาระยะห่างของทั้ง LSF ตัวแรกกับตัวกลาง และ LSF ตัวกลางกับตัวท้ายให้มีอัตราส่วนต่อระยะห่างของ LSF ที่อยู่ติดกันให้น้อยกว่า 0.7 การที่อัตราส่วนนี้สามารถมีค่าได้มากกว่าในกรณีของ LSF 2 ตัวเนื่องจากการเกิดการเรโซแนนซ์นั้นได้จากการช่วยเหลือกันของ LSF ทั้ง 3 ตัว

หลังจากหาตัวแทนที่น่าจะทำให้เกิดความถี่ฟอร์แมนท์ได้แล้ว จะทำการประมาณความถี่ฟอร์แมนท์ได้โดยการหาค่าเฉลี่ยของ LSF ทุกตัวที่ทำให้เกิดฟอร์แมนท์นี้ และผลลัพธ์ที่ได้จะต้องนำไปจัดกลุ่มในทางเวลาสำหรับแต่ละความถี่เพื่อขจัดความไม่ต่อเนื่อง โดยการค้นหาด้วย search window ในทางความถี่ที่ขึ้นกับเวลาสำหรับการจัดกลุ่มความถี่ของแต่ละเฟรมเข้าเป็นเส้นความถี่ฟอร์แมนท์เดียวกัน ค่าตัดสินใจต่างๆ ในการจัดกลุ่มนี้หาจากการทดสอบปรับหาจุดเหมาะสม โดยพิจารณาเทียบผลลัพธ์กับสเปคโตรแกรม เนื่องจากความถี่ฟอร์แมนท์ไม่สามารถหาค่าที่ถูกต้องแน่นอนได้ในเชิงวิเคราะห์ที่ทุกเวลาได้เพียงแบบเดียว จึงไม่สามารถทำ closed-loop optimization ค่าตัดสินใจข้างต้นได้ง่ายนัก



รูปที่ 4 การประมาณความถี่ฟอร์แมนท์ของคำว่า 'กิน' โดยระเบียบวิธีอย่างง่าย

5. สรุปและข้อเสนอแนะ

ผลการศึกษาลักษณะสมบัติของความถี่ LSP พบว่าช่วงความถี่ที่มีความถี่ LSP รวมกลุ่มกันอยู่หนาแน่น จะเป็นความถี่ฟอร์แมนท์ ของ vocal tract จากการทดสอบการหาความถี่ฟอร์แมนท์ด้วยการจัดกลุ่มความถี่ LSP เปรียบเทียบกับสเปคโตรแกรมและผลตอบเชิงความถี่ที่แท้จริงของ LSP พบว่าได้ผลที่ดี และเมื่อสังเกตโดยละเอียดพบว่าปัจจัยที่มีผลต่อการเกิดความถี่ฟอร์แมนท์มีสองประการคือความใกล้ชิดของคู่ความถี่ LSP และอัตราส่วนของคู่ความถี่นี้กับคู่ความถี่ใกล้เคียง จึงสามารถนำผลที่ได้นี้ไปสร้างเป็นระเบียบวิธีอย่างง่ายซึ่งสามารถประมาณความถี่ฟอร์แมนท์ได้อย่างใกล้เคียง แต่ผลที่ได้ต้องนำไปทำการประมวลผลเชิงเวลาเพื่อกำจัดความไม่ต่อเนื่องและผลจากสัญญาณรบกวนออก

ข้อควรปรับปรุงของระเบียบวิธีนี้คือ ผลลัพธ์สำหรับเสียงในช่วงที่เป็น unvoiced ยังไม่ดีนัก ซึ่งเป็นผลของการใช้แบบจำลองแบบ LP ในเครื่องลงรหัส G.729 ควรจะมีการปรับปรุงแก้ไขการประมาณค่าความถี่ฟอร์แมนท์ในช่วงดังกล่าว โดยนำค่า pitch period ซึ่งจะถูกส่งมาด้วยมาร่วมในการตัดสินใจด้วย

เอกสารอ้างอิง

[1] ITU-T RECOMMENDATION G.729, "Coding of Speech at 8 kbit/s using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP)," 1996

[2] เกรียงศักดิ์ ศิวะสนธิวัฒน์, "การศึกษาการรู้จำเสียงพูดโดยตรงจากเครื่องเข้ารหัส G.729," รายงานปริญญานิพนธ์ ภาควิชาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหา วิทยาลัย ปีการศึกษา 2539 หน้า 4-1 ถึง 4-4

[3] L. Rabiner and B-H. Juang, ***Fundamental of Speech Recognition***, Prentice-Hall, 1993, pp. 267-270

[4] J. Deller, J. Proakis, J. Hansen, ***Discrete-Time Processing of Speech Signals***, Macmillan, 1993, pp. 331-332

ประวัติผู้เขียนวิทยานิพนธ์

นายสิริ วงศ์วรชาติกาล เกิดเมื่อวันที่ 24 กรกฎาคม พ.ศ. 2519 ที่กรุงเทพมหานครจบการศึกษาระดับปริญญาวิศวกรรมศาสตรบัณฑิต ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ปี พ.ศ.2540 เข้าศึกษาต่อในหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปี พ.ศ.2540