NECTEC
Technical Journal
ISSN 1513-2145
วารสารวิชาการเนคเทค ปีที่ 2 ฉบับที่ 7 เดือน มีนาคม – มิถุนายน 2543 Vol. II, No. 7, March – June, 2000

Editor

Tutorial

Short Paper

Full Paper

Abstract

ECTI-21 NECTEC
National Electronics and Computer Technology Center
National Science and Technology Development Agency
สวทช NSTDA

Editor's Note

First of all, thank you to all readers, authors, and reviewers of NECTEC Technical Journal (NTJ) Vol. I. Time has passed so fast. This is the first issue of NTJ for the second year. There are a lot of changes in this new NTJ such as a new format, a new cover's page, a new publication's period, a new size, and a new subscription's fee (but not a new editor). In this year, the paper will be published for every four-month with doubling in size compared with NTJ of the first year.

From your feedback, we discovered that there are many Thai researchers either live in Thailand or study abroad like to write technical papers and share their discoveries. So what NTJ can do is to publish these quality technical papers and distribute to all interested IT readers. Meanwhile, as always, keep those emails coming in if you have a suggestion for us.

It's been hectic for the past few months for the NTJ editorial team. We've just been very busy with the preparation of proceedings of NECTEC Annual Conference 2000 : ECTI Technologies and the New Economy which took place at the United Nation Conference Center on June 22-25, 2000. There are 52 technical papers published in the proceedings. Abstract of each paper are extracted and published here. If anyone is interested to obtain full papers, please feel free to contact the Multimedia Technology Service at 644-8150…9 ext. 725 for further information. In this year conference, there was a contest to select three best papers in the category of best technical, best presentation, and most impact to Thai society. These three papers are reprinted in this issue of NTJ for your instant access.

The editorial team would like to express our appreciation to Dr. Thaweesak Koanantakool, director of NECTEC, for his vision and support in letting us continue our work for NTJ. Not to be outdone, NTJ is undergoing an upgrade. Please do email me your comments, compliments and - best of all - criticisms!

Sincerely,

Chularat Tanprasert
Editor of NTJ

ระบบโทรศัพท์เคลื่อนที่รุ่นที่ 3 (IMT-2000): มาตรฐาน หลักการทำงาน และเทคโนโลยี*

The Third Generation Mobile Phones (IMT-2000): Standards, Principles and Technologies

มัลลิกา อุณหวิวรรธน์ ผศ.ดร.สินชัย กมลภิวงศ์ สุธน แซ่หว่อง

ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์

อ.หาดใหญ่ จ. สงขลา 90112

ABSTRACT -- In this paper, we present an investigation of standards, principles, and technologies of the 3rd generation mobile phone, known as IMT-2000 (International Mobile Telecommunications 2000). It is expected that IMT-2000 mobile phones will offer broad-range of services to users; which include high quality audio, real-time video, video conferencing, facsimile, information retrieval such as file transfer and Internet access. The third-generation radio air interface standards, proposed by US, Japan, Europe, and Korea, based on CDMA (code division multiple access) are given. Introductory of IMT-2000, CDMA characteristics, frequency allocation, and basic frame structure of FDD (frequency division duplex) are presented. The evolution and migration of second generation to third generation mobile phones will be discussed.

Keywords -- IMT-2000, CDMA, 3rd Generation Mobile Phones, FDD

บทคัดย่อ -- ในบทความนี้ เป็นการศึกษาค้นคว้าและรวบรวมข้อมูล เกี่ยวกับมาตรฐาน,โครงสร้างพื้นฐาน และเทคโนโลยี ของระบบโทรศัพท์เคลื่อนที่รุ่นที่ 3 ซึ่งรู้จักในชื่อ IMT-2000 (International Mobile Telecommunications 2000) โทรศัพท์เคลื่อนในรุ่นที่ 3 เป็นระบบที่สามารถให้บริการพาหุสื่อแบบเคลื่อนที่และข้อมูลความเร็วสูง สิ่งนี้นำไปสู่ความต้องการทางด้านเทคนิคของ IMT-2000 องค์กรจากประเทศต่างๆทั้งอเมริกา ยุโรป จีน ญี่ปุ่น และเกาหลี ได้เลือกการเชื่อมต่อแบบ CDMA (code division multiple access) ในการจัดทำมาตรฐาน นอกจากนี้ในบทความยังกล่าวถึง การจัดสรรความถี่ที่ใช้ใน IMT-2000 ภาพรวมของ CDMA และโครงสร้างพื้นฐานของการสื่อสารข้อมูลแบบ FDD (frequency division duplex) การพัฒนาระบบในรุ่นที2 ไปสู่ระบบในรุ่นที่ 3 และการออกแบบระบบในรุ่นที่ 3 ถูกกล่าวโดยย่อ ในตอนท้ายของบทความ

คำสำคัญ: ไอเอ็มที-2000 ซีดีเอ็มเอ ระบบโทรศัพท์เคลื่อนที่รุ่นที่ 3 เอฟดีดี

1. บทนำ

ระบบโทรศัพท์เคลื่อนที่ในรุ่นที่2 (Second Generation Mobile Phones) ที่ใช้กันอยู่ปัจจุบัน ทั่วโลกได้แก่ระบบ GSM (Global System for Mobile communication) D-AMPS (Digital-Advanced Mobile Phone Systems) PDC(Personal Digital Cellular standard) และ IS-95 (Interim Standard-95) ซึ่งเป็นระบบที่ใช้เทคโนโลยีดิจิตอล (Digital Technology) เป็นพื้นฐานทำให้ระบบมีประสิทธิภาพสูงและประสบความสำเร็จในการจัดการบริการข้อมูลต่างๆไปยังผู้ใช้ได้ดีกว่าระบบที่ 1 ที่ใช้เทคโนโลยี อนาล็อก (AnalogTechnology)[1][2] จากการสำรวจของUMTS Forum (Universal Mobile Telecom-munications System Forum) เกี่ยวกับการเพิ่มขึ้นของจำนวนผู้ใช้บริการโทรศัพท์เคลื่อนที่ พบว่า ในปี 2000 จะมีผู้ใช้บริการ 400 ล้านคน และจะเพิ่มขึ้นเป็น 1,700 ล้านคน ในปี 2010 โดยเฉพาะอย่างยิ่งในแถบภาคพื้นเอเชียจะมีอัตราเพิ่มขึ้นมากกว่าภาคพื้นแถบอื่นๆ และในแถบยุโรปเปอร์เซ็นต์ของทราฟฟิกในการให้บริการทางด้านพาหุ

* สนับสนุนโดยทุนโครงการวิจัยโทรศัพท์เคลื่อนที่รุ่นที่ 3 (IMT-2000), NECTEC 2542

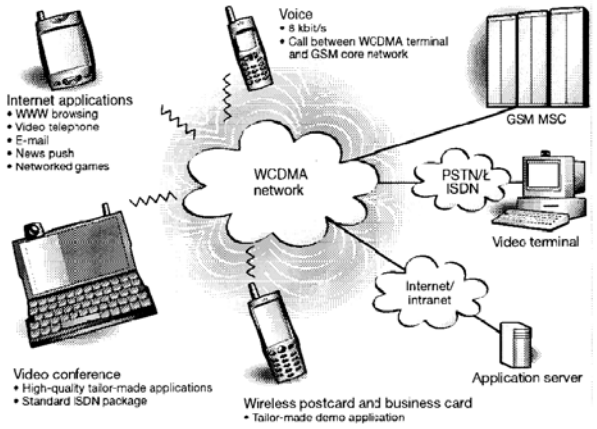
สื่อในโทรศัพท์เคลื่อนที่เพิ่มขึ้น 60% ในปี 2010 [3][4] เมื่อจำนวนของผู้ใช้บริการเพิ่มมากขึ้น จึงต้องมีการพัฒนาระบบโทรศัพท์เคลื่อนที่ให้ดียิ่งขึ้นเพื่อรองรับความต้องการของผู้ใช้บริการ ที่ต้องการให้โทรศัพท์เคลื่อนที่ไม่ใช่เป็นเพียงอุปกรณ์ที่ใช้สื่อสารสำหรับเสียงเพียงอย่างเดียว [5] เนื่องจากระบบโทรศัพท์เคลื่อนที่รุ่นที่ 2 เป็นการสื่อสารข้อมูลด้วยความเร็วต่ำ ทำให้ไม่สามารถรองรับความต้องการของได้ จึงทำให้เกิดเป็นแรงผลักดันให้เป็น " ระบบโทรศัพท์เคลื่อนที่ในรุ่นที่ 3 (Third Generation Mobile Phones)" ในโครงการ IMT-2000 (International Mobile Telecommunications 2000) ขององค์กรโทรคมนาคมระหว่างประเทศ (ITU: International Telecommunica-tions Union) โดยมีจุดมุ่งหมาย เพื่อให้มีการจัดทำมาตรฐานทางด้านการเข้าถึงของคลื่นวิทยุ เพื่อให้ระบบต่างๆ ที่กำลังใช้งานอยู่ในขณะนี้ สามารถถูกรวมเข้าด้วยกันเป็นมาตรฐานเดียวกัน ทำให้ผู้ใช้บริการไม่รู้สึกถึงการเปลี่ยนแปลงเมื่อมีการเคลื่อนที่จากระบบหนึ่งไปอีกระบบหนึ่ง อีกทั้งยังให้บริการได้ครอบคลุมได้ทุกพื้นที่ไม่ว่าจะเป็นพื้นที่แบบใดก็ตามเช่นภูเขา ชนบทหรือในเมืองที่มีประชากรอาศัยอย่างหนาแน่นและสามารถให้บริการหลายๆอย่างในเวลาเดียวกัน

การใช้สเปกตรัมให้มีประสิทธิภาพและคุ้มค่ามากที่สุดจึงเป็นหัวข้อสำคัญสำหรับการสร้างระบบโทรศัพท์เคลื่อนที่ในรุ่นที่ 3 เพื่อตอบสนองความต้องการของผู้ใช้บริการ ระบบในรุ่นที่ 3 ใช้พื้นฐานการเชื่อมถึงแบบCDMA (Code Division Multiple Access) ในเทคนิคของ FDD (Frequency Division Duplex) และ TDD (Time Division Duplex) เพื่อรองรับการส่งข้อมูลทั้งแบบ circuit switch และ packet-oriented switch [6] บทความนี้กล่าวถึง เป้าหมายพื้นฐานของ IMT-2000 ซึ่งอยู่ในหัวข้อที่ 2 และในหัวข้อถัดมาจะกล่าวถึงการจัดสรรความถี่เพื่อใช้ใน IMT-2000 สำหรับหัวข้อที่ 4 กล่าวถึงการดำเนินการในการจัดทำมาตรฐานขององค์กรต่างๆทั่วโลก ต่อมาในหัวข้อที่ 5 กล่าวถึงโครงสร้างทางเทคนิคของข้อเสนอที่อยู่ใน 3GPP (The Third Generation Partner Project) หัวข้อที่ 6 กล่าวถึงการพัฒนาไปยังระบบโทรศัพท์เคลื่อนที่รุ่นที่ 3 โดยมีการจำลองระบบที่จะนำมาใช้ในอนาคต หัวข้อสุดท้ายเป็นบทสรุปของบทความ

2. เป้าหมายพื้นฐานใน IMT-2000

ระบบในรุ่นที่3 เดิมใช้ชื่อโครงการว่า FPLMTS (Future Public Land Mobile Telecommunication Systems) แต่เนื่องจากชื่อเดิมนั้นมีความยาวไม่สะดวกในการเรียกชื่อ จึงได้เปลี่ยนชื่อใหม่เป็นคำว่า "IMT-2000" โดยคาดหวังให้ระบบโทรศัพท์เคลื่อนที่รุ่นที่ 3 นี้ สามารถให้บริการได้หลังจากปี 2000 และ ใช้งานในแถบความถี่ช่วง 2000 MHz โดยมีจุดมุ่งหมายดังนี้ [7][8][9]

- สามารถรองรับอัตราความเร็วของข้อมูลได้หลายระดับคือ 2 Mbps สำหรับการเคลื่อนที่อย่างช้าๆเช่น ในสำนักงาน อัตราความเร็ว 384 kbps สำหรับผู้ใช้ที่เคลื่อนที่ภายนอกอาคาร และ อัตราความเร็ว 144 kbps สำหรับผู้ใช้ที่เคลื่อนที่ด้วยความเร็วสูงเช่นขับยานพาหนะ
- เป็นระบบที่ไร้รอยต่อ (Global Roaming) คือจะไม่เกิดปัญหาเมื่อมีการเปลี่ยนจากระบบหนึ่งไปยังอีกระบบหนึ่ง ไม่ว่าผู้ใช้จะเดินทางไปที่ใดก็ตาม ก็ยังคงสามารถใช้โทรศัพท์ได้ทุกสถานที่ทุกเวลา (anywhere – anytime)
- สามารถให้บริการพาหุสื่อแบบเคลื่อนที่ได้เช่น ข้อความ (Text) ภาพนิ่ง (Image) และ วีดีโอ (Video) ให้บริการทางด้านอินเตอร์เน็ต (Internet) ไปรษณีย์อิเล็กทรอนิคส์ (e-mail)การประชุมทางวิดิทัศน์ (video conference) เป็นต้น
- ระบบสามารถให้บริการหลายๆบริการได้พร้อมกันในเวลาเดียวกัน คือเสียง ข้อมูลและพาหุสื่อ อีกทั้งระบบจะต้องมีความยืดหยุ่นสูงเพื่อรองรับกับการบริการใหม่ที่จะเกิดขึ้นในภายหลัง
- อุปกรณ์มีน้ำหนักเบา และ ราคาไม่แพง การให้บริการในระบบ IMT-2000 แสดงในรูปที่ 1 [32]



รูปที่ 1 การให้บริการในระบบ IMT-2000

ITU-R (ITU-Radio Communication Standardization Sector) จึงให้แต่ละประเทศส่งข้อเสนอ (proposal) สำหรับการพัฒนามาตรฐานของเทคโนโลยีสื่อสัญญาณทางด้านคลื่นวิทยุ IMT-2000 RTT – Radio Transmission Technology ต่อมา ITU-R ได้รับข้อเสนอเป็นจำนวนทั้งหมด 15 ข้อเสนอซึ่งแบ่งเป็นระบบ terrestrial 10 ข้อเสนอ และระบบ satellite 5 ข้อเสนอ สำหรับข้อเสนอแบบ terrestrial ใช้เทคโนโลยีในการเข้าถึงข้อมูลสามารถแบ่งได้ 2 รูปแบบคือ TDMA (Time Division Multiple Access) และ CDMA (Code Division Multiple Access) สำหรับข้อเสนอที่ใช้แบบTDMAเป็นการพัฒนาจากระบบเดิมเพื่อสามารถให้บริการในระบบรุ่นที่ 3 ได้ ส่วนข้อเสนอที่ใช้แบบ CDMA

สามารถแบ่งเป็นกลุ่มใหญ่คือ cdma 2000 และ WCDMA (Wideband-CDMA)[10]

WCDMA มีความหมายคือ ในระบบเดิมที่ใช้พื้นฐานของCDMAจะมีความกว้างของแถบความถี่เท่ากับ 1.25 MHz เมื่อมีการพัฒนาไปสู่ระบบในรุ่นที่ 3 จึงเพิ่มความกว้างของแถบความถี่เป็น 5MHz ซึ่งทำให้ระบบมีประสิทธิภาพสูงขึ้นในประเทศต่างๆได้มีการจัดตั้งระบบโทรศัพท์เคลื่อนที่รุ่นที่ 3 ขึ้น เช่นประเทศแถบยุโรปได้จัดตั้งระบบที่ชื่อว่า UMTS (Universal Mobile Telecommunications System) โดยใช้เทคโนโลยีทางด้าน WCDMA ซึ่งมีมาตรฐานที่กำหนดอยู่ใน IMT-2000 เพื่อให้เกิดความมั่นใจได้ว่าไม่ว่าผู้ใช้บริการจะเดินทางไปที่ใดก็สามารถใช้โทรศัพท์ของตนได้ทั่วโลกโดยไม่ต้องกังวลกับระบบที่แตกต่างกัน

3. การจัดสรรความถี่สำหรับ IMT-2000

ในด้านเศรษฐกิจ การใช้สเปกตรัมให้คุ้มค่ามีความจำเป็นมากสำหรับระบบใหม่ จึงได้มีการจัดประชุมเพื่อจัดสรรความถี่ขึ้น มีชื่อว่า WARC-World Administrative Radio Conference 1992 โดยแบ่งการใช้ช่วงของความถี่ออกเป็น 2 ช่วงคือ ช่วง 1,885-2,025 MHz และ 2,110-2,200 MHz ดังรูปที่ 2 [3] จะเห็นได้ว่าในทุกๆประเทศ มีการใช้ความถี่ในช่วงความถี่ต่ำอยู่บ้างแล้ว จึงไม่มีช่วงของความถี่ส่วนรวมที่สามารถใช้ได้ทั่วโลกสำหรับ IMT-2000 ดังนั้น ITU จึงต้องมีการจัดประชุมอีกครั้งเพื่อกำหนดการใช้แถบความถี่ที่เหมาะสมเพิ่มเติม ซึ่งได้จากการประชุม WARC 2000 [3][11]

4. การดำเนินการขององค์กรในการจัดทำมาตรฐาน IMT-2000

ในการจัดทำมาตรฐานสำหรับระบบในรุ่นที่ 3 นั้น ยังคงต้องคำนึงถึงการเข้ากัน (compatibility) ได้ระหว่างเทคโนโลยีเดิมกับเทคโนโลยีของระบบในรุ่นที่ 3 เพื่อไม่ให้เกิดการกระทบที่เสียหายกับระบบที่กำลังให้บริการอยู่ โดย ITU-R ได้มีการจัดตั้งทีมงานที่มีชื่อว่า Task Group 8/1 (TG8/1) ซึ่งทำหน้าที่ในการจัดทำข้อกำหนดในด้านเทคโนโลยีสื่อญญาณทางด้านคลื่นวิทยุนี้ องค์กรของประเทศต่างๆ ที่จัดทำระบบ สำหรับ IMT-2000 มีดังนี้

- European Telecommunications Standards Institute (ETSI) Special Mobile Group (SMG) ในภาคพื้นยุโรป
- Research Institute of Telecommunications Transmission (RITT) ในประเทศจีน
- Association of Radio Industry and Business (ARIB), Telecommunication Technology Committee (TTC) ในประเทศญี่ปุ่น
- Telecommunications Technologies Association (TTA) ในเกาหลี
- Telecommunications Industry Association (TIA) และ Telecommunications Planning Group (T1P1) ในภาคพื้นอเมริกา

ETSI SMG ของยุโรป ได้ทำการส่งข้อเสนอที่ชื่อว่า UTRA (UMTS Terrestrial Radio Access) โดยใช้เทคโนโลยีของ WCDMA และ TD-CDMAสำหรับการเชื่อมต่อของระบบ UMTS โดย WCDMA จะใช้เทคนิคของระบบการสื่อสาร 2 ทางแบบ FDD และ TD-CDMA ใช้เทคนิคแบบ TDD จึงทำให้เครื่องโทรศัพท์ในระบบ UMTS มีความยืดหยุ่นในการใช้งาน การให้บริการได้ครบถ้วน และ สามารถใช้กับระบบเดิมที่มีอยู่คือ GSM ได้ [12]

สำหรับโครงสร้างมาตรฐานของญี่ปุ่น ARIB ได้ส่งข้อเสนอที่ชื่อว่า WCDMA โดยข้อเสนอของญี่ปุ่นนี้จะมีความคล้ายคลึงกับของทางด้านยุโรป โดยใช้เทคนิคแบบ FDD เป็นหลัก และคาดว่า ญี่ปุ่นจะเป็นประเทศแรกที่จะนำระบบโทรศัพท์เคลื่อนที่รุ่นที่ 3 มาใช้จริงในปี 2001[13]

ในภาคพื้นอเมริกา ได้ส่งข้อเสนออยู่หลายฉบับคือ UWC-136 (Universal Wireless Communications-136) เป็นข้อเสนอที่พัฒนามาจาก IS-136 จากระบบในรุ่นที่ 2 WIMS (Wireless Multimedia and Messaging Services) เป็นข้อเสนอที่ใช้เทคโนโลยี WCDMA และ cdma 2000 เป็นข้อเสนอที่พัฒนาจาก IS-95 ซึ่งทั้ง 3 ข้อเสนอนี้ เป็นการจัดทำโดยกลุ่ม TIA และอีก 1 ข้อเสนอเป็นของ T1P1 มีชื่อว่า NA:WCDMA (North American: Wideband CDMA) ซึ่งข้อเสนอนี้มีความคล้ายคลึงกับของ UTRA ต่อมา NA:WCDMA และ WIMS ได้มีการรวมกันเป็นข้อเสนอเดียวแล้วเปลี่ยนชื่อเป็น WP-CDMA (Wideband Packet CDMA)[14][15]

TTA ของเกาหลี ได้ส่งข้อเสนอ 2 ข้อเสนอ สำหรับ IMT-2000 คือ CDMA I และ CDMA II โดยฉบับหนึ่งจะมีโครงสร้างที่ใกล้เคียงกับ ARIB W-CDMA ของญี่ปุ่น และอีกฉบับหนึ่งก็มีโครงสร้างที่ใกล้เคียงกับ cdma 2000 ของอเมริกา [16]

ในประเทศจีน ได้มีส่งข้อเสนอไปยัง ITU-R มีชื่อว่า TD-SCDMA (Time Division-Synchronous Code Division Multiple Access) โดยใช้พื้นฐานของ TD-CDMA แบบ ซิงโครนัส (synchronous) เพื่อใช้ใน TDD และ Wireless Local Loop (WLL)

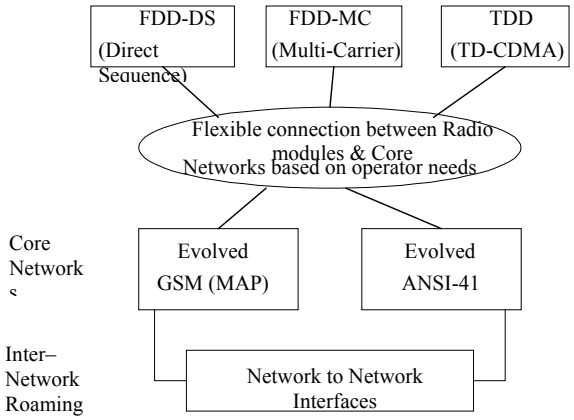
ข้อเสนอเหล่านี้ถูกส่งไปยัง ITU-R เพื่อจัดทำให้เกิดความสอดคล้องกัน เป็นการนำไปสู่ความเป็นมาตรฐานเดียวกันทั่วโลกซึ่งจะทำให้เกิดการใช้ประโยชน์อย่างคุ้มค่า ไม่ว่าจะเป็นผู้ใช้บริการ ผู้ลงทุนในการสร้างระบบ

โครงข่าย และผู้ผลิตในขั้นแรก การรวมรวบข้อเสนอเหล่านี้ให้เป็นมาตรฐานเดียวกันจะมีการแบ่งเป็น 2 โครงการ คือ

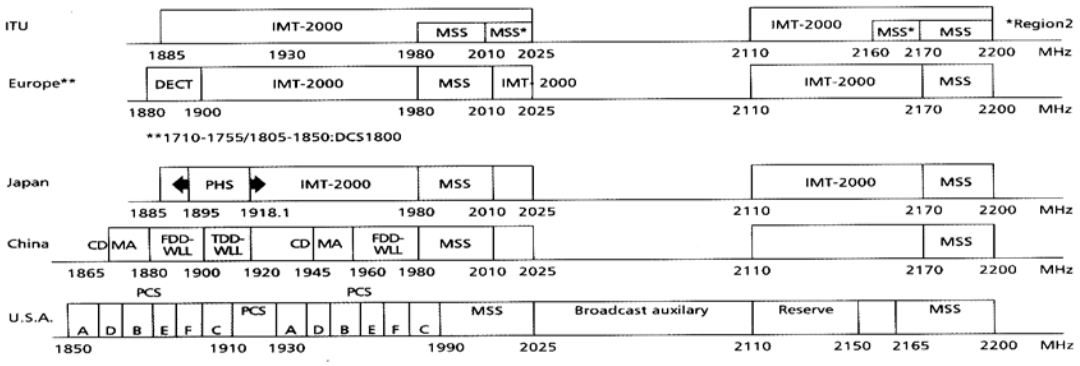
- 3GPP – The Third Generation Partnership Project เป็นโครงการที่จัดทำมาตรฐาน สำหรับข้อเสนอ ที่ใช้ WCDMA ซึ่งได้แก่ ETSI ARIB TTC TTA และ T1[17]
- 3GPP2 สำหรับข้อเสนอที่ใช้ พื้นฐานของ cdma 2000 ได้แก่ TIA และ TTA[18]

การทำงานของ 3GPP และ 3GPP2 นั้นอยู่ภายใต้การสนับสนุนของ Operator Harmonization Group (OHG) และจากโครงการ 3GPP และ 3GPP2 ก็จะต้องนำมารวมกันอีกครั้งเพื่อจัดทำเป็นมาตรฐานเดียวกันแล้วนำเชื่อมโยงกับ corenetwork

ANSI-41 และ GSM MAP สุดท้ายแล้วระบบต่างๆเหล่านี้จะถูกเชื่อมต่อเป็นโครงข่ายเดียวกันเป็นระบบโทรศัพท์เคลื่อนที่รุ่นที่ 3 ดังรูปที่ 3[19] [20][21]



รูปที่ 3การเชื่อมโยงโครงข่ายในIMT-2000



รูปที่ 2 การจัดสรรความถี่ทั่วโลก

5. ข้อเสนอทางด้านเทคนิค สำหรับ IMT-2000 ใน 3GPP

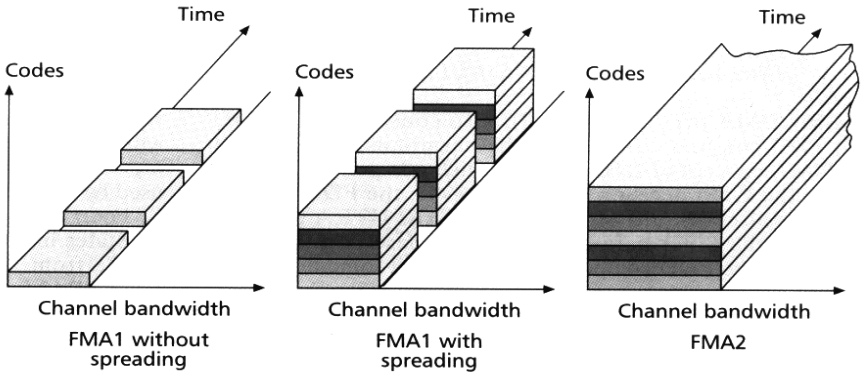
ความกว้างของแถบความถี่ในWCDMA เพิ่มขึ้นจากระบบเดิมมาเป็น 5 MHz เพื่อใช้ในระบบรุ่นที่3 เพราะ

- ความกว้างแถบความถี่ 5 MHz ทำให้สามารถรองรับการบริการข้อมูลที่ความเร็ว 144 และ 384 kbps ตามที่ต้องการได้เป็นอย่างดี
- ความกว้างที่ 5MHz นี้ทำให้แก้ไขปัญหาการ การกระทบของคลื่นได้ดีกว้างแถบความถี่ที่แคบกว่า ทำให้มีการเพิ่มไดเวอร์ซิตี้ (diversity) และ ปรับปรุงสมรรถนะ (performance)
- จากระบบเดิมซึ่งมีความกว้างของแถบความถี่ที่แคบกว่าทำให้ในการรับข้อมูลจะมีการขาดเป็นช่วงๆ เมื่อมีการเพิ่มแถบความถี่ให้กว้างขึ้นทำให้อัตราการรับที่ขาดเป็นช่วงๆนั้นจะลดลงจากเดิมและหากมีการเพิ่มความกว้างแถบความถี่ได้มากขึ้นเป็น 10 15 20 MHz ก็จะสามารถรองรับการให้บริการข้อมูลที่มีความเร็วสูงขึ้นกว่านี้ได้และระบบก็จะมีประสิทธิภาพมากขึ้น [22]

5.1 ยุโรป: พื้นฐานของ UTRA สำหรับ 3GPP

เนื่องจากรูปแบบ การเชื่อมต่อที่ใช้กันอยู่ในระบบเดิมนั้น มีความแตกต่างกันระหว่าง TDMA และ CDMA ดังนั้นทางแถบยุโรปจึงได้มีการค้นคว้าวิจัย การเชื่อมต่อ ขึ้นในระบบU MTS เพื่อให้เกิดความสอดคล้องกันระหว่าง TDMA และ CDMA ซึ่งมีชื่อโครงการว่า FRAMES (Future Radio Wideband Multiple Access System) ซึ่งเมื่อนำมาใช้ในการเข้าถึงข้อมูล มีชื่อเรียกว่า FMA (FRAMES Multiple Access) สามารถแบ่งได้เป็น 2 โหมด ดังรูปที่ 4 ดังนี้

- FMA 1 : เป็นการเข้าถึงข้อมูลแบบ wideband TDMA
- FMA 2 : เป็นการเข้าถึงข้อมูลแบบ wideband CDMA



รูปที่ 4 พื้นฐานการเข้าถึงข้อมูลแบบFMA นำมาใช้ใน UTRA

FMA1 with spreading จะใช้รองรับเทคนิคแบบ TDD เพราะ TDD เป็นการผสมระหว่าง TDMA กับ CDMA ทำให้ในการเข้าถึงข้อมูลต้องมีการแบ่งเป็นช่องเวลา (time slot) และมีการใช้รหัสเพื่อทำให้มีช่องสัญญาณเพิ่มมากขึ้น ดูรูปประกอบได้ในรูปที่ 4 ส่วน FMA 2 จะใช้เมื่อมีการเข้าถึงแบบ FDD เพราะไม่มีการใช้ช่องเวลาในการส่ง สำหรับ FMA1 without spreading เหมาะกับข้อเสนอของ UWC-136 ซึ่งใช้พื้นฐานของ TDMA

การใช้เทคนิค FDD และ TDD ทำให้สามารถรองรับความต้องการในการให้บริการที่มีความแตกต่างกันในระบบโทรศัพท์รุ่นที่ 3 ได้อย่างมีประสิทธิภาพ และรองรับการทำงานของการให้บริการข้อมูลแบบเคลื่อนที่ เช่นการเข้าสู่ อินเตอร์เน็ต การทำเทเลแบงกิ้ง (telebanking) ในเครื่องโทรศัพท์เคลื่อนที่

ในบางครั้งอาจกล่าวได้โดยรวมว่า UTRA เป็นข้อเสนอแบบ WCDMA เพราะจะใช้หลักของ WCDMA มากกว่า TD-CDMA พารามิเตอร์ของ UTRA สำหรับระบบโทรศัพท์เคลื่อนที่รุ่นที่ 3 แสดงในตารางที่ 1

5.2 ญี่ปุ่น: พื้นฐานของ WCDMA สำหรับ 3GPP

ข้อเสนอ WCDMA ของ ARIB นั้นใช้เทคนิคของ FDD และ TDD เทคนิค FDD ของWCDMA มีลักษณะคล้ายกับ UTRA ของยุโรป เพราะในปี 1998 ทั้งสองได้มีการประชุมเพื่อจัดวางรูปแบบร่วมกันและให้บริการแบบสมมาตร(symmetric) ของ uplink (เป็นการสื่อสารจากสถานีเคลื่อนที่ไปยังสถานีฐาน) และ downlink (เป็นการ สื่อสารจากสถานีฐานไปยังสถานีเคลื่อนที่) ส่วน TDD นั้นถูกออกแบบโดยมีโครงสร้างที่คล้ายกับFDD และให้บริการแบบไม่สมมาตร (asymmetric) ของ uplink และ downlink ดังนั้นส่วนใหญ่เมื่อกล่าวถึง WCDMA จึงมักหมายถึง WCDMA ของ ETSI/ARIB พารามิเตอร์ของ WCDMA สำหรับระบบโทรศัพท์รุ่นที่ 3 แสดงใน ตารางที่ 1 [13][22][23][24]

5.3 โครงสร้างพื้นฐานของภาคพื้นอเมริกา สำหรับ 3GPP

เมื่อเปรียบเทียบกับพื้นที่แถบอื่นๆแล้ว ในภาคพื้นแถบนี้มีการใช้ระบบโทรศัพท์เคลื่อนที่ รุ่นที่ 2 อยู่ด้วยกันหลายระบบ ทำให้เมื่อมีการพัฒนาไปสู่ IMT-2000 จึงเกิดความต้องการที่หลากหลายเพื่อทำการพัฒนาให้เกิดการรวมเข้ากันได้ของทั้ง 2 ระบบ ปัจจุบันการจัดตั้ง core network ในภาคพื้นอเมริกา เพื่อรองรับการเชื่อมถึงของระบบเดิม มีอยู่ด้วยกัน 2 เครือข่าย คือ ANSI–41 และ GSM MAP

ANSI–41 ในระบบรุ่นที่ 2 เป็น core network ที่ใช้กับระบบ IS-95 IS136 และ AMPS ดังนั้นในการจัดมาตรฐาน และคุณสมบัติของ ANSI-41 จึงเป็นหน้าที่ของคณะกรรมการใน TIA และเมื่อมีการพัฒนาไปสู่ IMT-2000 การดำเนินการพัฒนาจึงเป็นหน้าที่ของกลุ่มเทคนิคของ 3GPP2 ANSI-41 จึงเป็น core network ที่ใช้สำหรับ cdma 2000

GSM MAP ในระบบรุ่นที่ 2 เป็น core network ที่ใช้ในระบบ GSM ในการจัดทำมาตรฐานของ GSM MAP จึงเป็นหน้าที่ของคณะกรรมการใน ETSI SMGจากยุโรป และ T1P1.5 ในอเมริกา เมื่อมีการพัฒนาไปสู่ IMT-2000 ทำให้การพัฒนาGSM MAPถูกดำเนินการโดย 3GPP และ ให้สอดคล้องกับความต้องการของภาคพื้นอเมริกาใน T1P1

ตารางที่ 1 พารามิเตอร์พื้นฐานของFDD และ TDD ใน UTRA และ ใน WCDMA ARIB

	โครงสร้าง UTRA ของ ETSI SMG		โครงสร้าง WCDMA ของARIB	
	UTRA FDD	**UTRA FDD**	**WDMA FDD**	**WDMA TDD**
ระบบการจัดช่องสัญญาณ	WCDMA (DS-CDMA)	TD-CDMA	WCDMA (DS-CDMA)	TD-CDMA
อัตราเร็วชิป (chip rate)	4.096 Mchips/s	4.096 Mchips/s	4.0961.024/8.192/1 6.384)Mchips/s	4.0961.024/8.192/16.384) Mchs/s
ความกว้างของแถบความถี่	5 MHz	5 MHz	5 MHz	5 MHz
ความยาวกรอบ	10 ms	10 ms	10 ms	10 ms
วิธีการมอดูเลทข้อมูล	QPSK	QPSK	QPSK	QPSK
วิธีการมอดูเลท spreading downlink / uplink	QPSK	QPSK	QPSK/HPSK	QPSK
ระยะห่างของช่องเวลา	625 us	625 us	625 us	625 us
รูปร่างของพัลซ์	0.22	0.22	0.22	0.22

การส่งข้อเสนอสำหรับ IMT-2000 ไปยัง ITU-R จากภาคพื้นอเมริกานั้นมีอยู่ด้วยกัน 4 ข้อเสนอคือWCDMA-NA จาก T1P1 WIMS จาก TR 46.1 (TIA) UWC จาก TR45.3 (TIA) และ cdma 2000 จาก TR45.5 (TIA) ต่อมา WCDMA-NA และ WIMS ได้รวมกันเป็น WP-CDMA และจัดส่งไปยัง ITU-R WP-CDMA จึงเป็นอีกข้อเสนอหนึ่งที่ใช้เทคโนโลยีของ WCDMA เมื่อ T1 ส่งข้อเสนอไปยัง ITU-R จึงนับได้ว่าอเมริกาก็เป็นอีกหนึ่งในสมาชิกของ 3GPPอเมริกาไม่เพียงแต่เป็นหนึ่งสมาชิกของ 3GPP เท่านั้น อีกด้านหนึ่งการทำงานของ 3GPP2 และ UWCC ได้ทำงานไปพร้อมๆกันด้วยเพื่อพัฒนาเทคโนโลยีสำหรับ IMT-2000 ดังนั้นสำหรับในภาคพื้นอเมริกาแล้วมาตรฐานของ IMT-2000 โดยใช้ CDMA จะไม่มีเพียงแค่มาตรฐานเดียวในภาคพื้นแถบนี้ นอกจากนี้ผู้ดำเนินการในการจัดตั้งระบบสำหรับระบบโทรศัพท์เคลื่อนที่รุ่นที่ 3 ในภาคพื้นอเมริกายังแบ่งเป็น 2 กลุ่ม คือ กลุ่มที่ไม่ต้องการนำเทคโนโลยีที่ใช้ในระบบเดิมมาใช้ใน ระบบที่ 3 โดยกลุ่มนี้จะใช้การเชื่อมถึงแบบ WCDMA เราเรียกกลุ่มนี้ว่า greenfield operators และอีกกลุ่มหนึ่ง เป็นกลุ่มที่ต้องการพัฒนาเทคโนโลยีจากระบบเดิมที่ใช้งานกันอยู่ให้สามารถนำไปใช้ในระบบใหม่ได้และยังสามารถใช้กับแถบความถี่เดิม เรียกกลุ่มนี้ว่า established operators ดังนั้นความต้องการที่นำระบบGSM และ UWC-136 มาพัฒนาไปสู่ IMT-2000 จึงเป็นความต้องการของ established operators โดยใช้เทคโนโลยี EDGE (Enhanced Data rate for Global Evolution) มาเป็นมาตรฐานให้กับทั้ง2ระบบ

EDGE เป็นเทคโนโลยีที่นำการมอดูเลทแบบ 8 PSK (Phase Shift Keying) มาใช้ทำให้สามารถให้บริการข้อมูลความเร็วสูงได้คือ สามารถให้บริการที่อัตราเร็ว 384 kbps สำหรับการเคลื่อนที่ภายนอกอาคาร ถึงแม้ว่า EDGE ไม่สามารถให้การบริการที่ความเร็ว 2 Mbps ตามความต้องการของITU-R ได้ แต่ผู้ดำเนินการในอเมริกามองว่าด้วยอัตราเร็ว 384 kbps ก็มีความเพียงพอต่อความต้องการทางการตลาดในอเมริกาแล้วเพราะระบบเดิมนั้น มีอัตราเร็วของข้อมูลเพียง 9.6 kbpsเท่านั้น และยังสามารถให้บริการทางพาหุสื่อที่ต้องการได้ ซึ่งดำเนินการบนแถบความถี่ที่มีอยู่เดิมจึงนับได้ว่าคุ้มค่ามากแล้ว อีกเหตุผลหนึ่งคือโครงสร้างของระบบโทรศัพท์รุ่นที่ 2 ได้มีการลงทุนไปเป็นจำนวนมากและคาดว่าจะต้องมีการลงทุนเพิ่มขึ้นเรื่อยๆจึงควรจะพัฒนาจากระบบเดิมไปสู่ระบบใหม่ EDGE ยังมีข้อดีอีกข้อหนึ่งคือ สามารถนำเข้ามาใช้งานในระบบเครือข่ายเดิม คือ GSM และ IS-136 ได้อย่างสะดวกและสามารถปรับเปลี่ยนได้ทีละเล็กละน้อย โดยไม่กระทบกระเทือน ก่อให้เกิดความเสียหายกับระบบเดิม [3][10][11][25]

6. โครงสร้างพื้นฐานของเทคนิคFDD และ TDD

เนื่องจากใน IMT-2000 มีการใช้เทคโนโลยีCDMA เป็นพื้นฐานในการเชื่อมต่อ ดังนั้นในขั้นแรกจะอธิบายถึงภาพรวมของCDMA

6.1 ลักษณะโดยรวมของ CDMA

CDMA เป็นการเข้าถึงข้อมูลโดยอนุญาตให้ผู้ใช้จำนวนมากสามารถส่งข้อมูลของตนไปในช่องความถี่เดียวกัน ณ เวลาเดียวกันได้ โดยที่ผู้ใช้แต่ละคนจะมีลักษณะที่แตกต่างกันก็คือ รหัส ซึ่งรหัสที่ใช้นี้จะประกอบด้วย Pseudo-Noise (PN) code และ orthogonal code หรือ บางครั้งเรียกว่า Walsh Code (WC) ซึ่งแต่ละ WC จะมีความยาว 64 บิท (bit) หรือ ชิป (chip) ดังนั้นในการกำหนดช่องสัญญาณที่ใช้เทคโนโลยีแบบ CDMA

จะใช้รหัสเป็นตัวกำหนดความแตกต่างของแต่ละช่องสัญญาณ ทำให้ในระบบโทรศัพท์เคลื่อนที่ที่ใช้เทคนิคของ CDMA ทำให้เซลล์ที่อยู่ข้างเคียงสามารถใช้ความถี่ที่เหมือนกันได้ [26][27][28]

6.2 TDD

เป็นการสื่อสารข้อมูลแบบ 2 ทาง ที่เกิดจากการรวมการทำงานของ TDMA และ CDMA ทั้ง uplink และ downlink จะใช้ความถี่เดียวกัน (unpaired band) โดยใช้ช่องเวลามาทำการซิงโครนัสกัน

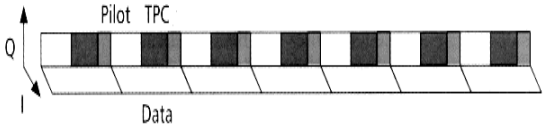
ทั้ง FDD และ TDD มีโครงสร้างของช่องสัญญาณที่เหมือนกัน สำหรับในบทความนี้จึงอ้างอิงของ FDD เป็นหลัก

6.3 FDD

เป็นการสื่อสารข้อมูลแบบ 2 ทาง คือ uplink และ downlink ในการส่งข้อมูล แบบ FDD จะใช้ความถี่ 1 คู่(paired band) เนื่องจากทั้ง uplink และ downlink จะไม่ใช้ความถี่เดียวกันในการส่ง โครงสร้างของช่องสัญญาณใน FDD มีความแตกต่างกันระหว่าง uplink และ uownlink คือ

6.3.1 ช่องสัญญาณของ uplink ในระดับชั้น physical

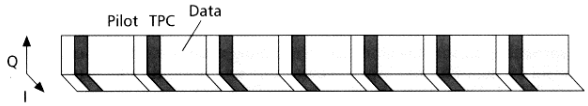
ช่องสัญญาณที่ใช้สำหรับส่งข้อมูลของผู้ใช้จะถูกส่งไปช่องสัญญาณเฉพาะของข้อมูล เรียกว่า DPDCH (Dedicated Physical Data Channel) และข้อมูลที่เกี่ยวกับการควบคุมถูกส่งไปในช่องสัญญาณเฉพาะของการควบคุม เรียกว่า DPCCH (Dedicated Physical Control Channel) ทั้ง DPDCH และ DPCCH เป็นแบบ IQ มัลติเพล็กซ์(multiplexed) คือ ใน DPDCH ข้อมูลจะส่งไปในแกน I และ DPCCH ซึ่งประกอบด้วยสัญลักษณ์ในการนำร่อง (Pilot symbols) สัญลักษณ์ในการควบคุมการส่ง (Transmitter Power Control symbols –TPC) จะถูกส่งไปในแกน Q ซึ่งแกน IQ จะมีความสัมพันธ์ทางด้านเฟส (phase) สามารถดูรูปประกอบได้ในรูปที่ 5 และมีการมอดูเลทแบบ QPSK (Quadrature Phase Shift Keying) จากที่ได้เคยกล่าวไว้แล้วว่าในเทคนิคแบบ CDMA เป็นการใช้รหัสในการแสดงความแตกต่างช่องสัญญาณ ดังนั้นการส่ง DPDCH และ DPCCH ไปยังคนละช่องสัญญาณแสดงว่า ทั้ง 2 ช่องสัญญาณก็ย่อมมีรหัสที่ต่างกัน คือ DPDCH จะมีรหัสเฉพาะเรียกว่า C_D และ DPCCH ก็จะมีรหัส C_C และเมื่อจะออกจากสถานีฐานก็จะถูกเข้ารหัสอีกครั้งเป็น C_{scramb} ซึ่งเป็นรหัสเฉพาะของแต่ละสถานีเคลื่อนที่ การทำ IQ มัลติเพล็กซ์โดยใช้รหัสนี้ ทำให้ในการส่งเป็นไปอย่างต่อเนื่องซึ่งการส่งอย่างต่อเนื่องนี้เองสามารถลดปัญหาการรวมตัวกับคลื่นแม่เหล็กไฟฟ้าได้ เพราะสถานีเคลื่อนที่มีโอกาสที่จะเข้าใกล้อุปกรณ์อิเล็กทรอนิคได้ง่าย ละการส่งแบบต่อเนื่องยังสามารถลดความต้องการเครื่องขยายในตัวสถานีเคลื่อนที่ได้



รูปที่5แสดงโครงสร้างช่องสัญญาณใน uplink

6.3.2 ช่องสัญญาณของ downlink ในในระดับชั้น physical

ในการมัลติเพล็กซ์ ของ DPDCH และ DPCCH จะเป็นการมัลติเพล็กซ์แบบเวลา(time multiplexed) ซึ่งดูรูปประกอบได้ในรูปที่ 6 ทั้ง DPDCH และ DPCCH ถูกส่งสลับกันไปตามเวลา ทำให้ทั้ง DPDCH และ DPCCH เป็นการส่งแบบไม่ต่อเนื่องซึ่งต่างจากใน uplink ในการส่งแบบใช้เวลามีข้อดีคือช่องสัญญาณควบคุมจะทำการค้นหาสถานีเคลื่อนที่ได้อย่างรวดเร็ว เนื่องจากการส่งแบบไม่ต่อเนื่องนั้น มีอัตราในการเสี่ยงต่อการเกิดการรวมตัวกับคลื่นแม่เหล็กไฟฟ้าแต่ใน downlink ปัญหาดังกล่าวจะมีโอกาสเกิดได้น้อยเพราะสถานีฐานแทบจะไม่มีโอกาสเข้าใกล้อุปกรณ์อิเล็กทรอนิค เหมือนกับสถานีเคลื่อนที่ในการเข้ารหัสของช่องสัญญาณใน downlink จะมีการเข้ารหัสดังนี้ ทั้ง DPDCH และ DPCCH จะมีรหัสที่เหมือนกันเรียกว่า C_{CH} และเมื่อมีการส่งสัญญาณไปยังสถานีเคลื่อนที่ก็จะมีการเข้ารหัสซึ่งเป็นรหัสเฉพาะสำหรับแต่ละสถานีฐาน คือ C_{scramb} [3][22][30][31]



รูปที่ 6 แสดงโครงสร้างช่องสัญญาณใน downlink

นอกจากในระดับชั้น physical ที่มีช่องสัญญาณแล้ว ในระดับชั้น logical ประกอบด้วยช่องสัญญาณอีกจำนวนหนึ่งซึ่งข้อมูลแต่ละชนิดที่จะถูกส่งไประหว่างสถานีเคลื่อนที่กับสถานีฐานจะใช้ช่องสัญญาณที่ไม่เหมือนกันขึ้นกับชนิดของข้อมูลนั้นๆ โครงสร้างของช่องสัญญาณ logical ใน WCDMA มีพื้นฐานตาม ITU-R M.1035 คือ

ก. ช่องสัญญาณ control (CCH) ประกอบด้วย

- ช่องสัญญาณ broadcast control (BCH) เป็นช่องสัญญาณที่ใช้ส่งข้อมูลที่เป็นคุณสมบัติของระบบและเซลล์
- ช่องสัญญาณ paging (PCH) ช่องสัญญาณนี้ ถูกใช้เมื่อเครือข่ายไม่ทราบว่าเซลล์ท้องถิ่นของสถานีฐาน
- ช่องสัญญาณ forward (FACH) เป็นช่องสัญญาณที่ส่งข้อมูลจากสถานีฐานไปยังสถานีเคลื่อนที่ ที่อยู่ภายในเซลล์เดียวกัน
- ช่องสัญญาณ dedicated control (DCCH) เป็นช่องสัญญาณที่ใช้ในการส่งข้อมูลในการควบคุมระหว่างเครือขายกับสถานีเคลื่อนที่

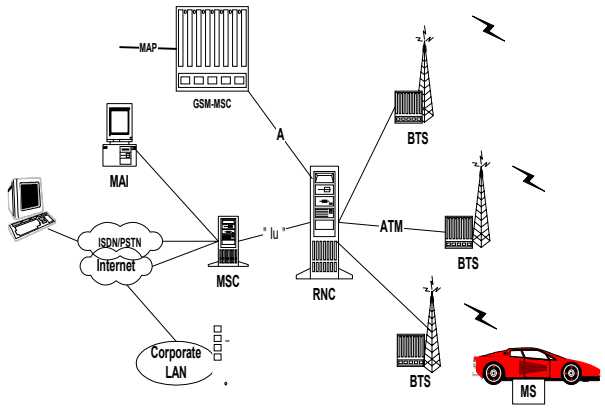
ข. ช่องสัญญาณ Traffic (TCH) ประกอบด้วย

- ช่องสัญญาณ dedicated traffic (DTCH) เป็น ช่องสัญญาณที่ส่งข้อมูลแบบจุดต่อจุดใน uplink และ downlink

- ช่องสัญญาณ user packet traffic (UPCH)

7. การพัฒนา ไปสู่ระบบโทรศัพท์เคลื่อนที่รุ่นที่ 3

การเปิดกว้างของระบบโทรศัพท์รุ่นที่ 2 ที่มีความแตกต่างกันในพื้นที่ทั่วโลก ให้สามารถพัฒนาไปสู่ ระบบที่ 3 ได้เริ่มขึ้นแล้ว อย่างเช่นในระบบ GSM 2+ ได้พัฒนาให้มีการบริการข้อมูลที่มีอัตราเร็วเพิ่มมากขึ้นได้และสามารถให้บริการทางด้านมัลติมีเดียได้บ้างแล้ว จากรูปที่ 7[6] เป็นภาพโดยรวมของระบบเครือข่าย WCDMA ซึ่งถูกออกแบบให้ใช้ใน ประเทศแถบทางยุโรปเพื่อรองรับ IMT-2000 ระบบจะประกอบด้วยชุมสายโทรศัพท์เคลื่อนที่ (Mobile Switching Center -MSC) อุปกรณ์ควบคุมเครือข่ายคลื่นวิทยุ (Radio Network Controller -RNC) สถานีฐาน (Base Transceiver Station –BTS) และสถานีเคลื่อนที่ (Mobile Station) เป็นเครื่องโทรศัพท์แบบ WCDMA สถานีฐานจะถูกออกแบบให้สามารถใช้ทรัพยากรส่วนต่างๆของระบบร่วมกันเพื่อประหยัดค่าใช้จ่ายและทำให้การใช้งานมีประสิทธิภาพ จากรูปที่ 7 สถานีฐานจะถูกเชื่อมต่อไปยังอุปกรณ์ควบคุมเครือข่ายคลื่นวิทยุ โดยใช้แพล็ตฟอร์มของ ATM (Asynchronous Transfer Mode) สำหรับใช้สื่อสารระหว่างโหนด(Node) ของเครือข่ายและภายในเครือข่ายแต่ละโหนดของ ATM ที่ใช้ใน IMT-2000 นี้จะใช้เทคโนโลยีสวิตซ์แพ็คเก็ตและ AAL2 (ATM Adaptation Layer 2) ได้ถูกพัฒนามาตรฐานขึ้นมาเพื่อรองรับการส่งของแพ็คเก็ตเสียง อุปกรณ์ควบคุมเครือข่ายคลื่นวิทยุของระบบจะทำหน้าที่ควบคุมเครือข่ายคลื่นวิทยุ เช่น การกำหนดและยกเลิกการเชื่อมต่อของช่องสัญญาณ การควบคุมกำลังส่ง (power control) การย้ายช่องสัญญาณ (handover) และฟังก์ชั่น codec ก็จะอยู่ภายในอุปกรณ์ควบคุมเครือข่ายคลื่นวิทยุ นอกจากนี้อุปกรณ์ควบคุมเครือข่ายคลื่นวิทยุ ยังมี A-interface เพื่อเชื่อมโยงไปยังชุมสายโทรศัพท์เคลื่อนที่ของเครือข่าย GSM ทำให้สามารถเชื่อมต่อช่องสัญญาณเสียงระหว่างเครื่องโทรศัพท์เคลื่อนที่ WCDMA กับระบบGSM ได้ ซึ่งเป็นไปตามข้อกำหนดใน IMT-2000 ที่ต้องการให้ระบบในรุ่นที่ 2 และรุ่นที่ 3 สามารถรวมเข้ากันได้ ส่วนชุมสายโทรศัพท์เคลื่อนที่ (MSC) ของระบบ มีหน้าที่หลักในการจัดตั้งการเชื่อมต่อช่องสัญญาณของเครื่องโทรศัพท์เคลื่อนที่ และมี Iu-interface สำหรับเชื่อมต่อกับอุปกรณ์ควบคุมเครือข่ายคลื่นวิทยุและสามารถเชื่อมต่อกับเครือข่ายแบบใช้สายเช่น ATM LAN ISDN และโมเด็ม อีกทั้งยังสามารถรองรับการบริการสื่อสารข้อมูลแบบ packet-switching และ circuit–switching นี่เป็นระบบจำลองที่จะนำมาใช้ใน IMT-2000 และยังคงต้องนำมาพัฒนาเพื่อให้ได้ระบบที่ดีที่สุดสำหรับใช้อนาคต แต่ไม่ว่าระบบโทรศัพท์เคลื่อนที่รุ่นที่ 3 จะเป็นอย่างไรก็ตาม การเข้ากันได้ระบบที่มีอยู่ก็เป็นอีกปัจจัยหนึ่งที่เป็นสิ่งสำคัญ



รูปที่ 7 ระบบเครือข่ายแบบ WCDMA ที่จะนำมาใช้ใน IMT-2000

8. บทสรุป

บทความนี้ได้นำเสนอภาพโดยรวมของ IMT 2000 ในส่วนของ โครงสร้างพื้นฐาน มาตรฐาน และเทคโนโลยีที่ใช้ใน IMT-2000 การศึกษาโครงสร้างพื้นฐานของ IMT-2000 เป็นอีกส่วนหนึ่งที่สำคัญในการสร้างและพัฒนาระบบโทรศัพท์เคลื่อนที่ รุ่นที่ 3 เพื่อให้ตอบสนองความต้องการของผู้ใช้บริการได้อย่างครบถ้วนดังนั้นองค์กรจากทั่วโลกทั้ง ยุโรป จีน ญี่ปุ่น เกาหลี และอเมริกา จึงได้มีการจัดทำข้อเสนอเพื่อมารองรับในการสร้างระบบโทรศัพท์เคลื่อนที่รุ่นที่ 3 โดยข้อเสนอจะเน้นไปสู่การพัฒนาด้าน air interface ซึ่งส่วนใหญ่จะเลือกใช้เทคโนโลยีแบบ WCDMA เป็นพื้นฐานในการพัฒนาระบบโดยใช้การสื่อสารข้อมูล 2 ทาง แบบ FDD เป็นหลักทำให้ระบบมีประสิทธิภาพที่สูงขึ้น สามารถรองรับการบริการข้อมูลแบบความเร็วสูงได้ข้อเสนอจากประเทศต่างๆถูกส่งไปยัง ITU-R เพื่อดำเนินการในการจัดทำมาตรฐานของข้อเนอเหล่านี้ให้เป็นมาตรฐานเดียวกัน ดังนั้น ITU จึงได้จัดตั้งกลุ่ม TG8/1 เพื่อดำเนินการจัดทำมาตรฐาน และคาดว่าการดำเนินการดังกล่าว จะทำให้ระบบสามารถให้บริการได้ทั่วโลกหลังปี 2000

เอกสารอ้างอิง

[1] N. Takezaki, ” The Next World: Standard for 3G Communications,” NTT DoCoMo’s IMT2000, November 1997
[2] W. Stark, ”Digital Communications Theory,” Technical paper no. 181, The University of Michigan, 2000
[3] P. Chaudhury, W.Mohr, and S.Onoe, ”The 3GPP Proposal for IMT-2000,” IEEE Communications Magazine, December 1999, pp. 72-81
[4] M. Callendar, ”IMT-2000 Standardization,” Telecom’99, 1999
[5] M. Gallagher ,W.Webb,”UMTS the next generation of mobile radio,” IEE Review, March 1999, pp.59-63
[6] P. Susampanpiboon, ” Third Generation (3G),” Internet Magazine, December1999, pp.68-75

[7] ITU Web page: http://www.itu.int/imt/1_infor/article/interview/index.html
[8] K. Etemad, "CDMA Concepts and Applications in Wireless PCS Networks," University of Maryland at College Park, October 1999
[9] R. M. Rajatheva, "Coding and Modulation Techniques for High Bit Rate Mobile Communication Systems: Third and Future Generation," The 1st Asia-Pacifi Seminar On Next Generation Mobiles Communications, January 2000
[10] P. Agrawal, D. Famolari, "Mobile Computing in Next Generation Wireless Networks," The 1st Asia-Pacific Seminar On Next Generation Mobiles Communications, January 2000
[11] P. Srisuksat, "The Migration from 2G to 3G for Thailand," The 1st Asia-Pacific Seminar On Next Generation Mobiles Communications, January 2000
[12] H. Benn, "ETSI SMG2 UTRA," ITU TG8/1 IMT-2000 Workshop, Jersey, November 1998
[13] ARIB W-CDMA, "Japan's Proposal for Candidate Radio Transmission Technology on IMT-2000: W-CDMA," ITU TG8/1 IMT-2000 Workshop, Jersey, November 1998
[14] TIA, "Evolution cdma-One to cdma 2000," ITU TG8/1 IMT-2000 Workshop, Jersey, November 1998
[15] UWC, "Evolution of TDMA to 3G," ITU TG8/1 IMT-2000 Workshop, Jersey, November 1998
[16] TTA , "Global CDMA I: Multiband Direct-Sequence CDMA System RTT Description," June 1998
[17] 3GPP Web page: http://www.3GPP.org
[18] 3GPP2 Web page: http://www.3GPP2.org
[19] M.Kijima, "Curent Status of IMT-2000 System Development," The 1st Asia-Pacific Seminar On Next Generation Mobiles Communications, January 2000.
[20] P.Weraarchakul, "CAT's cdma-ONe and the Migration towards 3G Coomunicatio Systems," The 1st Asia-Pacific Seminar On Next Generation Mobiles Communications, January 2000
[21] B.Chong, "Fujisu's Scope and Strategy for IMT-2000," The 1st Asia-Pacific Seminar On Next Generation Mobiles Communications, January 2000
[22] R.Prasad and T.Ojanpera', "An Overview of CDMA EvolutionToward Wideband CDMA," *IEEE Communications Surveys*, http://www.comsoc.org/pubs/surveys,Fourth Quarter 1998,Vol.1 No.1
[23] Qualcomm, "The Technical Case for Convergenced Third Generation Wireless Systems Based on CDMA", 1999
[24] A.Giordano and A.leverque, "Understanding Wireleless Communications ,CDMA and Next Generation Digital," Northeastern University,Octerber 1999
[25] P.Susampanpiboon, " Edge," Internet Magazine, December 1999, pp. 42-47
[26] T.Paugma, "Cellular Mobile Telephone System",1998
[27] H.Masaki, "IMT-2000 TDD Mode System," The 1st Asia-Pacific Seminar On Next Generation Mobiles Communications, January 2000
[28] Hewlett packard, "CDMA Overview & Testing," September 1999
[29] L.Harte, CDMA IS-95 for Cellular amd PCS, MCGraw-Hill Telecommunations,1999
[30] ARIB,"Specifications of Air-Interface for 3G Mobile System Ver.1.0," December 1997
[31] S.Barberis and E.Berruto," A CDMA-based radio into face for third Generation Mobile Systems," Mobile Networks and Applications, 1999, pp.19-29
[32] J.Eldstahl and A.Nasman, "WCDMA Evalition system-Evaluating the radio access technology of third-generation systems," Ericson Review, No.2,1999
[33] U.Black, "Mobile and Wireless Networks," Prentice Hall PTR,1996

คำย่อ

3GPP	The third Generation Partner Project
AAL	Asynchronous Transfer Mode Adaptation Layer
ARIB	Association of Radio Industry and Business
ATM	Asynchronous Transfer Mode
BCCH	Broadcast Control Channel
BTS	Base Transceiver Station
CCH	Control Channels
CDMA	Code Division Multiple Access
D-AMPS	Digital-Advanced Mobile PhoneSystems
DCCH	Dedicated Control Channel
DPCCH	Dedicated Physical Control Channel
DPDCH	Dedicated Physical Data Channel
EDGE	Enhanced Data rate for Global Evolution
ETSI SMG	European Telecommunications Standards Institute Special Mobile Group
FACH	Forward Access Channel
FDD	Frequency Division Duplex
FMA	Future Radio Wideband Multiple Access System Multiple Access
FPLMTS	Future Public Land Mobile Telecommunication Systems
FRAMES	Future Radio Wideband Multiple Access System
GSM	Global System for Mobile communication
IMT-2000	International Mobile Telecommunications - 2000
IS-95	Interim Standard-95
ISDN	Integrated Service Digital Network
ITU-R	International Telecommunications Union - Radio communication Standardization Sector

LAN	Local Area Network
MSC	Mobile Switching Center
NA:WCDMA	North American:Wideband Code Division Multiple Access
OHG	Operator Harmonization Group
PCH	Paging Channel
PDC	Personal Digital Cellular Standard
QPSK	Quadrature Phase Shift Keying
RITT	Research Institute of Telecommunications Transmission
RNC	Radio Network Controller
T1P1	Telecommunications Planning Group
TCH	Traffic Channels
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
TD-SCDMA	Time Division-Synchronous Code Division Multiple Access
TIA	Telecommunications Industry Association
TPC	Transmitter Power Control
TIA	Telecommunications Industry Association
TPC	Transmitter Power Control
TTA	Telecommunications Technologies Association
TTC	Telecommunication Technology Committee
UMTS	Universal Mobile Telecommunications System
UPCH	User Packet Traffic Channel
UTRA	Universal Mobile Telecommunications System Terrestrial Radio Access
UWC-136	Universal Wireless Communications-136
WARC	World Administrative Radio Conference
WCDMA	Wideband Code Division Multiple Access
WIMS	Wireless Multimedia and Messaging Services
WLL	Wireless Local Loop

Biographies

Dr. Sinchai KAMOLPHIWONG received the B.Sc.(EE) and M.E.(EE) degrees in Electrical Engineering from Prince of Songkla University (PSU), Thailand, in 1984 and 1988 respectively. In 1999, he received the Ph.D. (EE) degree from the University of NSW, Australia.

He was with the Department of Electrical Engineering, PSU, Thailand in 1984 where he tough in microprocessors and system designs. In 1992, he was with the Department of Computer Engineering at the same university where he involved in PABX and system control development projects. He is now an Assistant Professor in the Department of Computer Engineering, PSU. Thailand. He is a team leader of IP telephony development project. He is a research member of 3rd Generation Mobile Phone (IMT-2000) project. His main interest research areas are: flow control in ATM networks, IP networks and packet telephony, high speed networks, computer network protocols, network embedded systems, mobile networks, modelling and simulation in computer networks. He is a member of IEEE, ACM, ComSoc, and Computer.

มัลลิกา อุณหวิวรรธน์ สำเร็จการศึกษาจาก คณะวิศวกรรมศาสตร์ สาขาโทรคมนาคม ณ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ปีการศึกษา 2542. เป็นผู้ช่วยวิจัยโครงการพัฒนาโทรศัพท์เคลื่อนที่ยุคที่ 3 ปัจจุบันกำลังศึกษาปริญญาโท ภาควิชาเทคโนโลยีสารสนเทศ และวิศวกรรมไฟฟ้า คณะสถาบันเทคโนโลยีนานาชาติสิรินธร มหาวิทยาลัยธรรมศาสตร์ งานวิจัยที่สนใจ การเข้าถึงข้อมูลโดยการเข้ารหัส ระบบโครงข่ายเคลื่อนที่ไร้สาย

สุธน แซ่หว่อง วิศวกรรมศาสตร์บัณฑิต (เกียรตินิยมอันดับ 1) คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์ เมื่อปี พ.ศ. 2542 ปัจจุบันเป็นอาจารย์ประจำภาควิชาวิศวกรรมศาสตร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์ เป็นผู้ช่วยวิจัยโครงการพัฒนาโทรศัพท์เคลื่อนที่ยุคที่ 3 และเป็นผู้วิจัยหลักในโครงการพัฒนาระบบไอพีเทเลโฟนนี่ งานวิจัยที่สนใจ: ไอพีเทเลโฟนนี่ เครือข่ายพหุสื่อ การส่งข้อมูลแบบทันเวลาจริง

Progress Report on NECTEC's Microelectronics Project

Itti Rittaporn Chumnarn Punyasai and Pavan Siamchai

Thai Microelectronics Center (TMEC) Project
National Electronics and Computer Technology Center (NECTEC)
73/1 Rama VI Road, Ratchathevee, Bangkok 10400, Thailand
Phone (+66-2)644-8150 Fax. (+66-2)644-8137, E-Mail: itti@notes.nectec.or.th

ABSTRACT -- As a national center, NECTEC's road map in the area of microelectronics is to start from IC/VLSI design, wafer fabrication and then to expand into the areas of flat panel display devices, sensors, optoelectronics, and so on. The mission is to activate the rise of microelectronics industry in Thailand. After nearly ten years, however, the progress has not been significant. Main reason is the lack of strategy and commitment in the national policy-making level. Most of the activities are bottom-up oriented, and not well-enough supported. Under this situation, NECTEC is working hard to promote IC/VLSI design, and to realize Thailand's first wafer fabrication line. In this report NECTEC's microelectronics related activities are summarized.

1. Introduction

NECTEC has started its activity on micro-electronics in 1989 by joining the ASEAN - Australia cooperation project. Some VLSI have been designed in-house and sent for fabrication in Australia during the program.

In 1994, NECTEC has provided the software tools for VLSI design to about 10 universities in Thailand to support their educational program and activity on VLSI design. This has contributed to the gradual expansion of the design activities in Thailand and after several years, there are already some IC design companies starting their business.

In the same year, NECTEC signed an agreement with IMEC (Interuniversity Microelectronics Center) of Belgium under the agreement of cooperation in microelectronic sector between Thailand and Belgium governments. Under this agreement, NECTEC will send engineers and technicians for 50 man-months to IMEC for training in process technology and facilities management focusing, practically at 0.5 micron CMOS technology. IMEC will transfer basic process technology and provide necessary consultancy where Belgium government will subsidize the cost on IMEC's side. This agreement will end by 31 August 2000.

In 1995, the Cabinet has approved the proposal of NECTEC to establish the Thai Microelectronics Center (TMEC). The plan is to build a wafer fabrication line for CMOS process technology below 1 micron with the capacity of 500 six-inches wafers/month. The original objectives and goals of this project have been summarized in Ref.1. Under the proposal, the government will support 600 million Baht in 3 years for establishing the TMEC center and a private company, Alphatec Group, will donate another 300 million Baht for the project. Because they are planning to start wafer fabrication business in Thailand and need support in the area of human resource development.

2. Current Status and future plan

At present, NECTEC's microelectronics related activities can be summarized as follows.

2.1 TMEC Project

In 1997, NECTEC has started the construction of the TMEC building in the Alpha-technopolis Industrial Park at Chacherngsao province. Cleanroom, equipped with necessary facilities to support CMOS 0.5 micron technology, of class 100 (300 m^2) and class 1000 (700 m^2) are the main features. Originally the work was planned to be finished by March 1998, however due to the contractor financial problem after the economic crisis in 1997, the progress of construction had become very slow and out of schedule. Extensions of the contract have been made several times due to government measures to help construction industry from the economic crisis. This gave further delay to the construction and at present the contract is extended until 31 January 2000.

Figure 1 and 2 show the progress at the construction site as of August 1999. The amount of accomplished work is about 20% of the total.

Fig. 1 TMEC Construction site (August 99)

Fig. 2 TMEC Construction site (August 99)

Figure 3 shows the design drawing of the TMEC building.

Fig. 3 TMEC Building

For the process and metrology equipment, 10 items have been purchased. About 25 more are necessary.

For the TMEC Project as a whole, the bad economic situation from 1996 and the Baht crisis in 1997 have raised serious problems. Firstly, it was not possible for the Alphatec Group to donate the 300 million Baht to the project as plan. Secondly, with the only 600 million Baht from the government and the depreciation of the Baht after mid 1997, it is not possible to complete the project without additional financial support. From 1998 to mid 1999, a lot of effort has been spent to find way-out. Eventually the Project has been re-studied and the revised proposal of the TMEC Project has been approved by the NSTDA Board in August 1999. Now it is being submitted to the Minister of Science Technology and Environment to ask for the Cabinet approval. Only after that that NECTEC will have necessary funding to accomplish the plan of this project.

In the revised proposal, the main philosophy of the project has been changed from to be the center to mainly support human resource development for the wafer fabrication industry (process engineer), to the center to activate the rise of IC/VLSI design industry in Thailand (design engineer). This is considered more strategically appropriate since Thailand has a wide base of IC/VLSI design education network in more than 10 major universities with more are participating. A large pool of new designers can be expected to contribute to the rise of IC/VLSI design industry in Thailand in near future.

For the cooperation with IMEC, NECTEC has accomplished sending engineers for 50 man-months training to IMEC by August 1999. The training covered almost all aspects of wafer fabrication (process technology, process integration, design technology, fab management, and facilities design and maintenance, QA and safety, etc.) Under the agreement, IMEC will send experts to help commissioning TMEC's cleanroom and advising process start-up. Collaboration beyond the end of present agreement is under consideration.

2.2 MPC (5 um) Program with ERC/KMITL

Beginning in 1996, NECTEC has started a Multi Project Chip (MPC) program with Electronic Research Center (ERC) of King Mongkut's Institute of Technology Ladkrabang (KMITL) to upgrade the microelectronics facilities of ERC from 20 micron process to 5 micron CMOS process. The main aim is to make the facilities up-to-date and ready as a center of excellence for providing multi project chip fabrication service for universities over Thailand. A new cleanroom of class 1000 (30 m^2) and class 10,000 (111 m^2) has been built. LPCVD, ion implanter, etcher, furnace have been newly purchased or upgraded. After a long series of effort, it is planed to open its service by March 2000. Under this scheme, NECTEC and ERC/KMITL will arrange a MPC (Multi Project Chip) fabrication service for IC/VLSI designers.

Fig. 4 ERC/KMITL Cleanroom

2.3 IC/VLSI design and Promotion

As mentioned earlier, this activity is considered the most important at present because it is the mission to explore Thailand's new capability in IC/VLSI design area. The facts that it is not highly capital intensive (compare with wafer fabrication), the technology is still growing, the worldwide shortage of VLSI designers, and the chance of gaining business opportunity worldwide are convincing to most people. The success of this activity will be crucial for the rise of microelectronics industry in Thailand.

In-house design projects: At present , there are several in-house design projects running. First is a project to design the 8051 microcontroller compatible chip with higher performance. Another is the project to design a GPS chip (Global Positioning System).

This article is a reprint of a paper presented at the 1st NRCT-KOSEF (Thai-Korea) Workshop on Semiconductors Technology (Nov.1999, Bangkok)

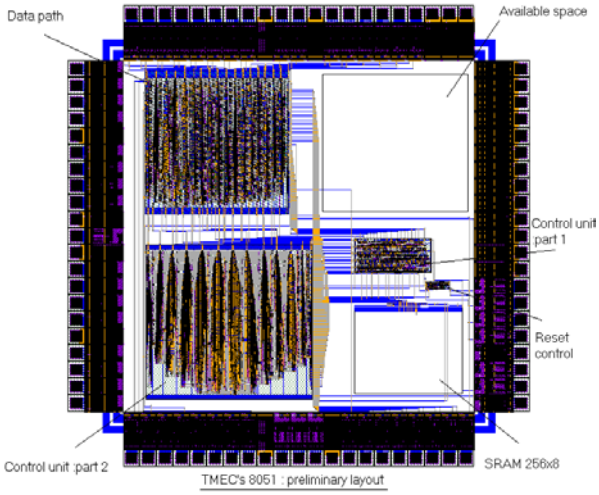


Fig. 5 TMEC's 8051 layout design

Promotion of IC/VLSI design To strengthen the design community, NECTEC has initiated several new activities.

1. **Establishing of Microelectronics Forum** In January 1999, NECTEC has initiated the microelectronics community in Thailand which comprises of universities, IC packaging industries, hard disk drive industries, system houses, and design houses to setup the Microelectronics Forum. The main objective is to utilize the expertise and resources of each party to organize training for industrial work force to help upgrading microelectronics industries in Thailand towards higher value added and upstream industry. The Forum is also aiming to cooperate in consolidation of appropriate policy/strategy and make suggestion to the Thai government.

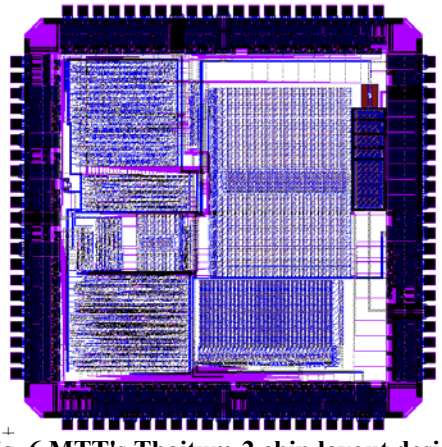


Fig. 6 MTT's Thaitum-2 chip layout design

2. **Establishing of IC Design Network** Also in January 1999, NECTEC has set up the IC Design Network. It is the network of people doing design in Thailand with members of about 40 at present. The main aims are to cooperate in both development of new IC/VLSI designers and sharing of expertise and IP resources in real design projects.

3. **Supporting private design houses:** Although still a few, there are already several design companies in Thailand. NECTEC is trying its best to support these companies. As an example, MTT (Microelectronic Technologies Thailand) is collaborating with NECTEC in submitting their designs for fabrication through IMEC of Belgium under EuroPractice Program. Figure 4 is the layout design of the "ThaiTum2" chip (microcontroller) of MTT. The company is also offering commercial service for layout design to outside customers.

4. **Development of AIT master program** AIT (Asian Institute of Technology) is planning to start its master degree in microelectronics from May 2000 academic year. NECTEC strongly involves in the establishment of this program and will actively support the program, providing faculties for some courses and also facilities for students' projects.

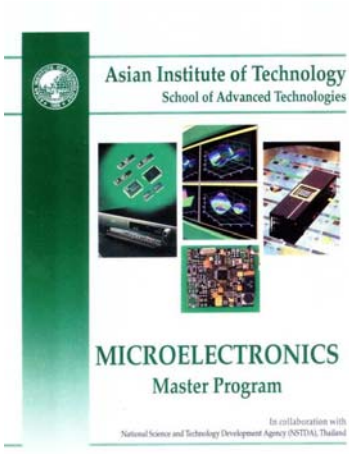


Fig. 7 AIT 's microelectronics program

5. **IC Design Contest 2000** As a result of the 3rd IC Design Network meeting held on 31 August 1999, a committee has been setup to run the Thailand first IC design contest to enhance the public knowledge and interest in IC/VLSI design. For this first "IC Design Contest 2000", specifications of the target chips, one analog and one digital, will be given and the chance is open to all designers. About 50 challengers are expected and the designs of the winners will be sent for real fabrication.

3. Summary

NECTEC as a national center is doing its best to promote microelectronics in Thailand. Activities span from academic cooperation on human resource development to promoting and supporting IC designers both in universities and private sectors. Without strong commitment and sufficient supports comparable to other countries, real strength and success of Thailand in microelectronics is not foreseeable.

4. Acknowledgment

The authors would like to thank all colleagues in the microelectronics area in Thailand; the IC Design Network members, colleagues in the TMEC project, ERC/KMITL and universities. Contribution of IMEC in our activities is highly appreciated.

References

[1] P. Sirirutchatapong, P. Sichanugrist, P. Siamchai, C. Punyasai, and S. Chinsakolthanakorn, "R&D Programs in fabrication of VLSI in Thailand", Proceedings of EECON-20, Nov. 13-14, 1997, Bangkok, Thailand

Itti Rittaporn received B.E. (1983), M.E.(1985) and Ph.D (1988) in applied physics engineering from the University of Tokyo. From 1988-1996 he joined Superconductivity Research Laboratory of Ministry of International Trade and Industry of Japan as senior researcher where he had been using photoelectron spectroscopy, STM/AFM/MFM for study of electronic structure of high-Tc superconductors. He returned to Thailand in 1996 under the mission of transferring the SORTEC synchrotron light source from Japan to Thailand. From 1997, He joined NECTEC and at present is heading TMEC Project.

Chumnarn Punyasai received B.Sc. in physics from Konkaen University in 1989 and received Master degree in computer engineering (VLSI design) from University of Southwestern Louisiana in 1993. He co-found NECTEC's Microelectronics Lab and is heading IC/VLSI Design Group of TMEC Project. He is at present working for Ph.D degree at KMITL.

Pavan Siamchai received B.E. (1990) and M.E.(1992) in electrical engineering from Chulalongkorn University. In 1995, he received his Ph.D. in electrical and electronics engineering from Tokyo Institute of Technology. His fields of expertise include thin film processing, amorphous silicon, silicon processing, semiconductor technology.

การค้นคืนสารสนเทศออนไลน์โดยใช้จีเนติกอัลกอริทึม

Online Information Retrieval using Genetic Algorithms

บังอร กลับบ้านเกาะ
สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ลาดกระบัง กรุงเทพฯ 10520
โทร (02) 7372551-4(EXT:802) โทรสาร 3269074 E-Mail:S0067034@kmitl.ac.th
เอื้อน ปิ่นเงิน
ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ลาดกระบัง กรุงเทพฯ 10520
โทร (02) 3269969 E-Mail:kpouen@kmitl.ac.th

ABSTRACT -- This article presents an online information retrieval using genetic algorithms to increase information retrieval efficiency. Under vector space model, information retrieval is based on the similarity measurement between query and documents. Documents with high similarity to query are judge more relevant to the query and should be retrieved first. Under genetic algorithms, each query is represented by a chromosome. These chromosomes feed into genetic operator process: selection, crossover, and mutation until we get an optimize query chromosome for document retrieval.

KEY WORDS -- GENETIC ALGORITHMS, INFORMATION RETRIEVAL

บทคัดย่อ -- บทความนี้นำเสนอเกี่ยวกับวิธีการค้นคืนสารสนเทศออนไลน์โดยประยุกต์ใช้จีเนติกอัลกอริทึมเพื่อเพิ่มประสิทธิภาพในการค้นคืนสารสนเทศ ภายใต้เวคเตอร์สเปซโมเดล (vector space model) การค้นคืนสารสนเทศใดนั้นขึ้นอยู่กับความคล้ายคลึง (similarity) ระหว่างเอกสารและคิวรี เอกสารใดที่มีความคล้ายคลึงกับคิวรีสูงย่อมแสดงว่าเอกสารนั้นมีความสัมพันธ์กับคิวรีมากกว่าและควรจะได้รับการค้นคืนขึ้นมาก่อน ในขั้นตอนของจีเนติกอัลกอริทึมนั้น คิวรีจะถูกแทนด้วยโครโมโซม ซึ่งโครโมโซมเหล่านี้จะถูกนำเข้าสู่กระบวนการจีเนติกโอเปอเรเตอร์ต่างๆ อันได้แก่ การคัดเลือก การครอสโอเวอร์ และการมิวเตชัน จนกระทั่งได้โครโมโซมคิวรีที่เหมาะสมเพื่อนำไปค้นคืนสารสนเทศต่อไป

คำสำคัญ -- จีเนติกอัลกอริทึม, การค้นคืนสารสนเทศ

1. คำนำ

จีเนติกอัลกอริทึมคิดค้นขึ้นโดย John Holland ในปีค.ศ. 1975 [1][2] เป็นการนำขบวนการวิวัฒนาการของสิ่งมีชีวิตมาประยุกต์ใช้ในงานปัญญาประดิษฐ์ เพื่อใช้สำหรับหาคำตอบที่ดีที่สุด (Optimization) ของปัญหาต่าง ๆ จากจำนวนคำตอบที่เป็นไปได้ทั้งหมดของการแก้ปัญหานั้น ปัจจุบันได้มีการประยุกต์ใช้จีเนติกอัลกอริทึม ในงานต่าง ๆ อย่างแพร่หลาย เช่น การใช้จีเนติกอัลกอริทึมในกับการแก้ปัญหาทางคณิตศาสตร์ การแก้ปัญหาการหาเส้นทางที่มีระยะทางสั้นที่สุด (Traveling Salesman Problem : TSP) [1] [2] การจัดตารางสอน (Timetable Scheduling Problem) [7] [12] การควบคุมหุ่นยนต์ (Robot Control) [8] [10] รวมทั้งการค้นคืนสารสนเทศด้วย [3] [4] [5]

2. งานวิจัยที่เกี่ยวข้อง

จีเนติกอัลกอริทึมเป็นกระบวนการแก้ปัญหาเอนกประสงค์ที่สามารถนำไปประยุกต์ใช้กับปัญหาประเภทต่าง ๆ ได้ สำหรับงานวิจัยที่เกี่ยวข้องเท่าที่ศึกษาและค้นคว้ามามีดังนี้ คือ “การเรียนรู้ของแขนหุ่นยนต์โดยการใช้การโปรแกรมพันธุการ” โดยจุมพล พลวิชัย [8], “การปรับปรุงประสิทธิผลของโปรแกรมหุ่นยนต์ซึ่งก่อกำเนิดโดยการโปรแกรมเชิงพันธุศาสตร์” โดย ธวัชชัย เอี่ยมมนัสกุล [10], “การลดทอนความเพียรพยายามเชิงคำนวณของวิธีการเรียนรู้แบบการโปรแกรมเชิงพันธุศาสตร์” โดยชัยวัฒน์ เจษฎาปกรณ์ [9], “การจัดตารางสอนของโรงเรียนแบบอัตโนมัติโดยจีเนติกอัลกอริทึม” โดยกาญจนี วงศ์วิภาพร [7] ในส่วนของการประยุกต์จีเนติกอัลกอริทึมกับการค้นคืนสารสนเทศ เช่น Jing-Jye Yang, Robert R. Korfhage และ Edie Rasmussen ได้ประยุกต์ใช้จีเนติกอัลกอริทึมในการปรับปรุงคิวรี (Query Improvement in Information Retrieval using Genetic Algorithms) [4] โดยมีวัตถุประสงค์ที่จะปรับปรุงคิวรี ให้สามารถค้นคืนข้อมูลได้ตรงตามความต้องการของผู้ใช้มากขึ้น ข้อจำกัดของงานวิจัยนี้คือใช้การปรับปรุงเฉพาะน้ำหนัก (term weight) ของคิวรี ไม่มีการเพิ่มเติมคำใหม่ ๆ เข้ามาในกระบวนการของ จีเนติก ฉะนั้นการค้นหาจะจำกัดอยู่เฉพาะที่คำที่มีอยู่ในคิวรีเดิม ในขณะที่ Maria J. Martin-Bustista, Henrik L. Larsen, Jacob Nicolaisen และ Torben Svendsen ได้ประยุกต์ใช้จีเนติกอัลกอริทึมร่วมกับยีนส์แบบฟัซซีในงานวิจัยเรื่อง An Approach to An Adaptive Information Retrieval Agent using Genetic Algorithms with Fuzzy Set Genes [5] โดยมีวัตถุประสงค์ที่จะสร้างเอเจนต์เพื่อการกลั่นกรองส่วนบุคคล (personal filter agent) ระหว่างผู้ใช้กับกลไกค้นหาข้อมูลบนอินเตอร์เน็ต (internet search engine) ข้อจำกัดของงานวิจัยนี้คือการใช้ยีนส์แบบฟัซซีทำให้ยากต่อการกำหนดรูปแบบโครโมโซมได้ถูกต้องยากแก่การใช้งานจริง

จะเห็นว่าแนวโน้มการใช้จีเนติกอัลกอริทึมในการค้นคืนสารสนเทศมีมากขึ้นและผลที่ได้ก็เป็นที่น่าพึงพอใจแต่ทั้งนี้ยังต้องมีการวิจัยพัฒนาต่อไป เพื่อเพิ่มประสิทธิภาพในการค้นคืนสารสนเทศให้ดียิ่งขึ้น

3. หลักการของจีเนติกอัลกอริทึม

3.1 องค์ประกอบของจีเนติกอัลกอริทึม

จีเนติกอัลกอริทึมมีองค์ประกอบที่สำคัญ 5
องค์ประกอบ คือ

1) รูปแบบโครโมโซมที่ใช้ในการนำเสนอทางเลือกที่สามารถจะเป็นได้ของแต่ละปัญหา
2) วิธีสร้างประชากรต้นกำเนิด (initial population) ของทางเลือกที่สามารถจะเป็นไปได้
3) ฟังก์ชันสำหรับประเมินค่าความเหมาะสม (fitness) เพื่อให้คะแนนแต่ละทางเลือก
4) จีเนติกโอเปอเรเตอร์ (Genetic Operator) ซึ่งใช้ในการปรับเปลี่ยนองค์ประกอบของข้อมูลตลอดกระบวนการ ได้แก่ การคัดเลือก การครอสโอเวอร์ และการมิวเตชัน
5) ค่าพารามิเตอร์ต่างๆ ซึ่งต้องใช้สำหรับ จีเนติกอัลกอริทึม เช่น ขนาดของประชากร, ความน่าจะเป็นของการใช้จีเนติกโอเปอเรเตอร์ และจำนวนรุ่นเป็นต้น
6) ฟังก์ชันสำหรับประเมินค่าความเหมาะสม (fitness) เพื่อให้คะแนนแต่ละทางเลือก
7) จีเนติกโอเปอเรเตอร์ (Genetic Operator) ซึ่งใช้ในการปรับเปลี่ยนองค์ประกอบของข้อมูลตลอดกระบวนการ ได้แก่ การคัดเลือก การครอสโอเวอร์ และการมิวเตชัน
8) ค่าพารามิเตอร์ต่างๆ ซึ่งต้องใช้สำหรับ จีเนติกอัลกอริทึม เช่น ขนาดของประชากร, ความน่าจะเป็นของการใช้จีเนติกโอเปอเรเตอร์ และจำนวนรุ่นเป็นต้น

3.2 การทำงานของจีเนติกอัลกอริทึม

ขั้นตอนแรกของจีเนติกอัลกอริทึมคือการกำหนดฟังก์ชันความเหมาะสม รวมทั้งรูปแบบโครโมโซมเสียก่อน จากนั้นจึงเริ่มสร้างประชากรต้นกำเนิดตามรูปแบบโครโมโซมที่ได้กำหนดไว้ เมื่อได้ประชากรต้นกำเนิดแล้วก็ทำการวัดค่าความเหมาะสม (fitness) ของแต่ละโครโมโซม เพื่อคัดเลือกเข้าสู่กระบวนการจีเนติกโอเปอเรเตอร์ โดยทำการคัดเลือกเอาเฉพาะโครโมโซมที่มีค่าความเหมาะสมเป็นที่น่าพอใจชุดหนึ่งเก็บไว้ โครโมโซมที่คัดเลือกไว้นั้นจะถูกนำมาทำการครอสโอเวอร์และมิวเตชันได้เป็นโครโมโซมชุดใหม่ ซึ่งเราจะนำโครโมโซมชุดใหม่นี้มาวัดค่าความเหมาะสมเพื่อทำการคัดเลือกและดำเนินการต่อไปจนสิ้นสุดตามเงื่อนไขที่ได้กำหนดไว้ ก็จะได้โครโมโซมที่มีค่าความเหมาะสมเป็นที่น่าพอใจ หรือได้คำตอบของปัญหาที่ต้องการดังแสดงในรูปที่ 1

4. การค้นคืนสารสนเทศออนไลน์โดยใช้จีเนติกอัล กอริทึม

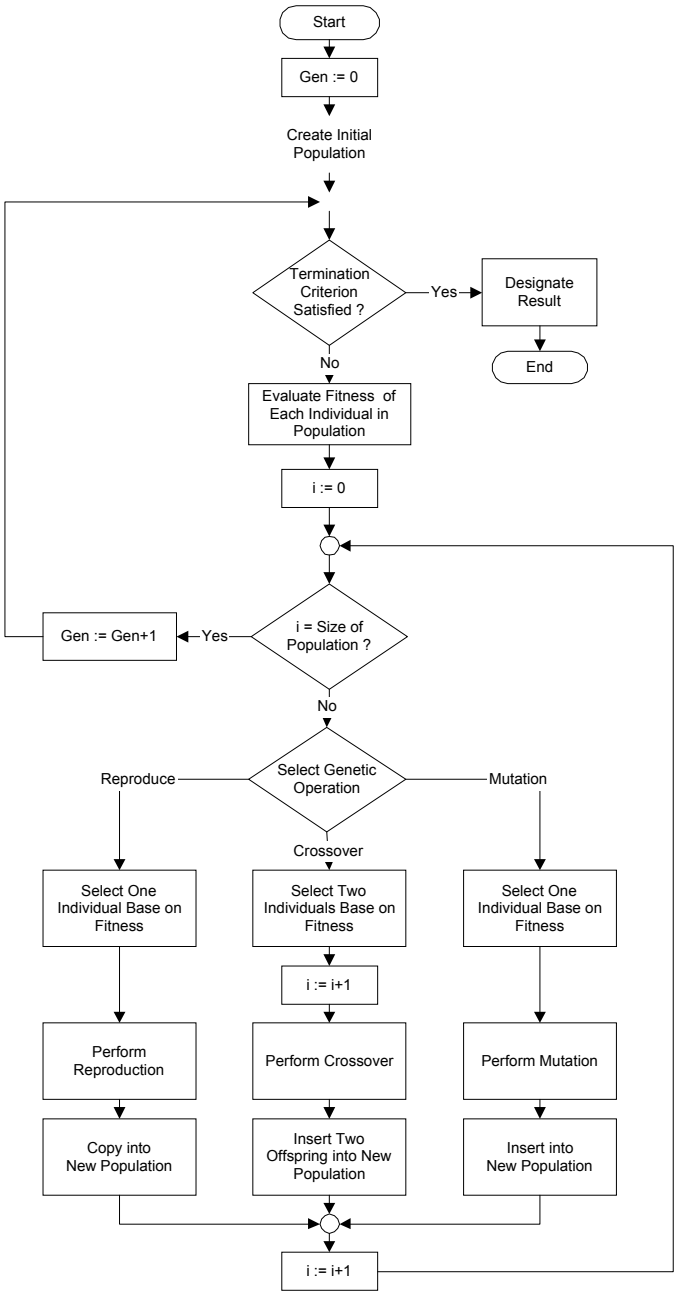
4.1 รูปแบบการนำเสนอ

ระบบค้นคืนสารสนเทศออนไลน์โดยใช้จีเนติกอัลกอริทึมนี้ จัดทำขึ้นภายใต้เวคเตอร์สเปซโมเดล โดยแต่ละเอกสารแทนด้วยเวคเตอร์ของคำสำคัญและคิวรีแทนด้วยเวคเตอร์ของคำที่ใช้ในคิวรี(query terms) การใช้เวคเตอร์คิวรีในการค้นคืนทำได้โดยการจับคู่ระหว่างเอกสารกับคิวรีแล้วทำการคำนวณหาค่าความคล้ายคลึง โดยถ้าหากปรากฏคำ ณ

ตำแหน่งนั้นในเอกสารหรือคิวรีจะให้ค่าเป็น "1" หากไม่ปรากฏให้ค่าเป็น "0" ค่าที่คำนวณได้เป็นการแสดงว่าเอกสารนั้น ๆ ตรงกับคิวรีเพียงใด ซึ่งค่าความคล้ายคลึงที่วัดได้นี้จะถูกนำไปใช้ในขั้นตอนการคัดเลือกของกระบวนการจีเนติก ตัวอย่างเช่น ให้ Doc เป็นเวคเตอร์ของเอกสาร และ Query เป็นเวคเตอร์ของคิวรี

$Doc = (term_1, term_2, \ldots term_n)$

$Query = (qterm_1, qterm_2, \ldots qterm_m)$



รูปที่ 1 การทำงานของจีเนติกอัลกอริทึม [12]

เมื่อผู้ใช้ใส่คิวรีเข้าสู่ระบบสามารถค้นคืนชุดของเอกสาร 5 ฉบับประกอบด้วยคำสำคัญดังนี้

Doc_1 ประกอบด้วย Relational Database, Query, Data Retrieval, Computer Networks, DBMS

Doc_2 ประกอบด้วย Artificial Intelligence, Internet, Indexing, Natural Language Processing

Doc_3 ประกอบด้วย Databases, Expert System, Information Retrieval System, Multimedia

Doc_4 ประกอบด้วย Fuzzy Logic, Neural Network, Computer Networks

Doc_5 ประกอบด้วย Object-Oriented, DBMS, Query, Indexing

สามารถจัดเรียงเป็นชุดของคำสำคัญ คือ Artificial Intelligence, Computer Networks, Data Retrieval, Databases, DBMS, Expert System, Fuzzy Logic, Indexing, Information Retrieval System, Internet, Multimedia, Natural Language Processing, Neural Network, Object-Oriented, Query, Relational Databases

นำเสนอในรูปแบบโครโมโซม ได้ดังนี้คือ

Doc_1 = 0110100000000011

Doc_2 = 1000000101010000

Doc_3 = 0001010010100000

Doc_4 = 0100001000001000

Doc_5 = 0000100100000110

โครโมโซมชุดแรกที่ได้มานี้จะเรียกว่าประชากรต้นกำเนิด ซึ่งจะนำไปผ่านกระบวนการจีเนติกต่อไป ความยาวของโครโมโซมเหล่านี้จะขึ้นอยู่กับจำนวนคำสำคัญของชุดเอกสารทั้งหมดที่ตรงตามคิวรี จากตัวอย่างแต่ละโครโมโซมมีความยาวเท่ากับ 16 บิท

4.2 การวัดค่าความเหมาะสม

ฟิตเนสฟังก์ชันหรือฟังก์ชันวัดความเหมาะสม คือ ฟังก์ชันที่ใช้ในการประเมินว่าแต่ละทางเลือก (solution) นั้น มีความเหมาะสม สามารถใช้แก้ปัญหาได้ดีเพียงใด สำหรับปัญหาของการค้นคืนสารสนเทศก็คือ ทำอย่างไรจึงจะสามารถค้นหาเอกสารที่ตรงตามความต้องการของผู้ใช้ จากการที่เลือกใช้เวคเตอร์สเปซโมเดล การที่จะกำหนดว่าเอกสารใดตรงตามคิวรีหรือคล้ายคลึงกับคิวรีนั้น สามารถทำได้โดยใช้ฟังก์ชันในการวัดความคล้ายคลึงซึ่งก็มีด้วยกันหลายรูปแบบ ในที่นี้เลือกใช้ฟังก์ชันดังตารางที่ 1 [6] เพื่อนำมาใช้เป็นฟิตเนสฟังก์ชัน

จากตารางที่ 1 กำหนดให้ $X=(x_1, x_2, x_3, .., x_n)$,

$|X|$ = จำนวนคำที่ปรากฏใน X ,

$|X \cap Y|$ = จำนวนคำที่ปรากฏทั้งใน X และ Y

ฟิตเนสฟังก์ชันมี 2 รูปแบบคือแบบถ่วงน้ำหนักและแบบไบนารี แต่ในการวิจัยครั้งนี้จะใช้แบบไบนารี

ผลลัพธ์ที่ได้จากการคำนวณของฟิตเนสฟังก์ชันจะมีค่าอยู่ระหว่าง 0.0–1.0 โดยที่ 1.0 หมายถึงเอกสารและคิวรีนั้นเหมือนกัน ค่าที่เข้าใกล้ 1.0 แสดงว่าเอกสารและคิวรีนั้นมีความสัมพันธ์กันมาก ส่วนค่าที่เข้าใกล้ 0.0 แสดงว่าเอกสารและคิวรีนั้นมีความสัมพันธ์กันน้อย ค่าที่ได้จากการคำนวณนี้เรียกว่าค่าความเหมาะสม (fitness)

ตารางที่ 1 ฟิตเนสฟังก์ชัน

Similarity Measure	**Binary Term Vectors**	**Weighted Term Vectors**
Dice coefficient	$2\frac{\lvert X \cap Y \rvert}{\lvert X \rvert + \lvert Y \rvert}$	$\frac{2\sum_{i=1}^{t} x_i . y_i}{\sum_{i=1}^{t} x_i^2 + \sum_{i=1}^{t} y_i^2}$
Cosine coefficient	$\frac{\lvert X \cap Y \rvert}{\lvert X \rvert^{1/2} . \lvert Y \rvert^{1/2}}$	$\frac{\sum_{i=1}^{t} x_i . y_i}{\sqrt{\sum_{i=1}^{t} x_i^2 . \sum_{i=1}^{t} y_i^2}}$
Jaccard Coefficient	$\frac{\lvert X \cap Y \rvert}{\lvert X \rvert + \lvert Y \rvert - \lvert X \cap Y \rvert}$	$\frac{\sum_{i=1}^{t} x_i . y_i}{\sum_{i=1}^{t} x_i^2 + \sum_{i=1}^{t} y_i^2 - \sum_{i=1}^{t} x_i . y_i}$

4.3 การคัดเลือก

หลังจากได้ค่าความเหมาะสมของแต่ละโครโมโซมแล้ว ขั้นตอนต่อมาก็คือการคัดเลือกสายพันธุ์ (Selection) การคัดเลือกสายพันธุ์เป็นไปตามหลักการอยู่รอดของสิ่งที่เหมาะสมที่สุด (Survival of the fittest) โดยโครโมโซมที่มีค่าความเหมาะสมเป็นที่น่าพอใจก็จะได้รับการคัดเลือกไว้ ส่วนโครโมโซมที่มีค่าความเหมาะสมต่ำ ก็จะได้รับการคัดเลือกน้อยหรือไม่ได้รับการคัดเลือกเลย

4.4 การครอสโอเวอร์

การครอสโอเวอร์ (Crossover) คือ การนำโครโมโซม 2 โครโมโซมมาทำการตามขั้นตอนต่างๆ ซึ่งจะให้ค่าโครโมโซมใหม่ที่จะนำไปใช้ในการคัดเลือกครั้งถัดไป หรือเป็นการนำโครโมโซมสองโครโมโซมมาผสมกันเพื่อให้ได้ค่าโครโมโซมใหม่ขึ้นมานั่นเอง ในขั้นตอนนี้จะพยายามสร้างทางเลือกใหม่และปรับปรุงทางเลือกให้ดีขึ้นโดยการครอสโอเวอร์ ซึ่งจีเนติกอัลกอริทึมจะพยายามสร้างทางเลือกที่ดีขึ้นโดยการรวมลักษณะที่ดีของแต่ละโครโมโซมเข้าด้วยกัน โครโมโซมที่มีค่าความเหมาะสมสูงกว่ามักจะถูกเลือกมาครอสโอเวอร์บ่อยครั้งกว่าส่งผลให้มีโอกาสในการรอดไปยังรุ่นต่อ ๆ ไปมีมากกว่า

การครอสโอเวอร์สามารถทำได้หลายวิธีเช่น การครอสโอเวอร์หนึ่งตำแหน่ง (one point crossover) การครอสโอเวอร์สองตำแหน่ง (two point crossover) หรือการครอสโอเวอร์หลายตำแหน่ง (multiple point crossover)

การครอสโอเวอร์แบบหนึ่งตำแหน่ง โดย ครอสโอเวอร์ ณ ตำแหน่งที่ 8

101111110011101

100110011110000

ผลที่ได้หลังการครอสโอเวอร์ คือ

10111111***1110000***

10011001***0011101***

4.5 การมิวเตชัน

การมิวเตชัน (Mutation) เป็นลักษณะของการผ่าเหล่าคือการนำโครโมโซมเก่ามาสุ่มแก้ไขบางส่วนของโครโมโซม เช่น บิทบางบิทให้เปลี่ยนไป ทำให้ได้โครโมโซมใหม่ที่มีสายพันธุ์ต่างจากเดิม ซึ่งมีโอกาสที่จะเป็นโครโมโซมที่ดีขึ้นหรือเลวลงก็ได้ หากโครโมโซมที่ได้ใหม่นี้เป็นโครโมโซมที่เลวลง โครโมโซมที่ได้นี้ก็จะถูกคัดออกไปในขั้นตอนการคัดเลือกเอง วัตถุประสงค์ของการมิวเตชันคือเพื่อประกันการสูญหายของข้อมูลและเพื่อความหลากหลายของข้อมูล ตัวอย่างของการทำมิวเตชัน เช่น สุ่มเลือกเปลี่ยนโครโมโซมตำแหน่งที่ 10

10111111***0***11101

ผลที่ได้คือ 10111111***1***11101

จีเนติกอัลกอริทึม จะทำเป็นวัฏจักรหมุนเวียนอยู่เช่นนี้จนกระทั่งถึงจุดหนึ่งตามเงื่อนไข โดยอาจสิ้นสุดเมื่อถึงรุ่น (generation) ตามที่กำหนดหรือสิ้นสุดเมื่อพบคำตอบที่ดีที่สุดแล้วหรือถึงเธรชส์โฮลด์ (threshold) ตามที่ได้กำหนดไว้ล่วงหน้าแล้วนั้น

4.6 ขั้นตอนการทำงานของระบบ

1. ผู้ใช้ใส่คิวรีเข้าสู่ระบบ
2. นำคิวรีไปค้นหาจากรายการคำสำคัญ ทั้งหมดของระบบที่มีอยู่
3. นำชุดของเอกสารที่ตรงตามคิวรีมาแปลงเป็นโครโมโซม จะได้เป็นประชากรต้นกำเนิด
4. นำชุดโครโมโซม (ประชากรต้นกำเนิด) ที่ได้นี้เข้าสู่กระบวนการจีเนติกโอเปอเรเตอร์ อันได้แก่ การคัดเลือก การครอสโอเวอร์ และการมิวเตชัน ตามที่ได้อธิบายไว้แล้ว

   ดำเนินการตามข้อ 4 จนกระทั่งถึงรุ่นที่กำหนด จะได้โครโมโซมคิวรีที่เหมาะสม (optimize query chromosome) เพื่อค้นคืนสารสนเทศต่อไป
5. แปลงโครโมโซมคิวรีที่เหมาะสมนี้เป็นคิวรีเพื่อค้นคืนสารสนเทศจากฐานข้อมูล

5. การทดลอง

5.1 วิธีการทดลอง

การทดลองเป็นการทดลองค้นคืนจากคิวรี 21 คิวรี ซึ่งใช้ฟิตเนสฟังก์ชันหรือฟังก์ชันวัดความเหมาะสมที่แตกต่างกัน 3 ฟังก์ชันดังที่ได้กล่าวมาแล้ว คือ Jaccard Coefficient, Cosine Coefficient และ Dice Coefficient โดยที่แต่ละฟิตเนสฟังก์ชันนั้นจะทดสอบโดยใช้ค่าความน่าจะเป็นในการครอสโอเวอร์ 0.8 และ ค่าความน่าจะเป็นในการมิวเตชัน 0.01, 0.10 และ 0.30 จำนวนรุ่นสูงสุด (Max Generation) เท่ากับ 30 รุ่น เพื่อทดสอบเปรียบเทียบประสิทธิภาพ โดยประสิทธิภาพของการค้นคืนสารสนเทศวัดได้จากค่าความแม่นยำ (Precision) และค่าความระลึก (Recall)

ค่าความระลึก (R) เป็นอัตราส่วนของการค้นพบเอกสารที่ถูกต้องจากจำนวนเอกสารที่ถูกต้องทั้งหมด ดังสมการที่ 1 [3][6]

$$R = \frac{\text{จำนวนเอกสารที่ถูกต้องที่ค้นคืนได้}}{\text{จำนวนเอกสารที่ถูกต้องทั้งหมดในฐานข้อมูล}} \quad (1)$$

ค่าความแม่นยำ (P) เป็นอัตราส่วนของการค้นพบเอกสารที่ถูกต้องจากจำนวนเอกสารทั้งหมดที่ทำการค้นคืนมาได้ ดังสมการที่ 2 [3][6]

$$P = \frac{\text{จำนวนเอกสารที่ถูกต้องที่ค้นคืนได้}}{\text{จำนวนเอกสารทั้งหมดที่ค้นคืนออกมาได้}} \quad (2)$$

ฐานข้อมูลสำหรับงานวิจัยนี้ เป็นฐานข้อมูลโครงงานนักศึกษา คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง จำนวน 343 โครงงาน

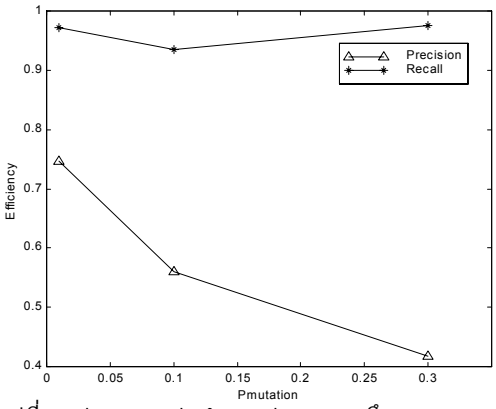
ตารางที่ 2 ผลการค้นคืนโดยใช้ฟิตเนสฟังก์ชันทั้ง 3 ฟังก์ชัน

Keywords	Query Chromosome	F1	F2	F3	RetRel	RetNRel
application	00100000000000000000000000001100	0.84	0.91	0.90	30	1
database	00010000000000000000000000000010000100	0.59	0.65	0.65	34	8
DNS	0011001001	1.00	1.00	1.00	6	2
internet	00000000000010000000000001	0.76	0.86	0.84	41	-
marketing	0110110	1.00	1.00	1.00	11	8
recognition	11000	0.71	0.75	0.74	7	-
security	000100100	1.00	1.00	1.00	17	57
network	00001000000000010000000	1.00	1.00	1.00	78	21

5.2 ผลการทดลอง

ปรากฏดังนี้คือ

1. ผลจากการทดลองโดยใช้ฟิตเนสฟังก์ชันทั้ง 3 นั้นทำให้ได้โครโมโซมคิวรีซึ่งเหมือนกัน เพียงแต่มีค่าความเหมาะสม (F1,F2,F3) ต่างกันเท่านั้น ดังนั้นจึงสามารถค้นคืนได้ชุดเอกสารชุดเดียวกัน แสดงผลได้ดังตารางที่ 2 โดย RetRel หมายถึงจำนวนเอกสารที่ถูกต้องที่ค้นคืนได้ และ RetNRel หมายถึงจำนวนเอกสารที่ไม่ถูกต้องที่ค้นคืนได้
2. ค่าความแม่นยำ การค้นคืนสารสนเทศโดยใช้ค่าความน่าจะเป็นในการครอสโอเวอร์ 0.8 และค่าความน่าจะเป็นในการมิวเตชัน 0.01 จะให้ค่าความแม่นยำสูงสุดคือ 0.746 ในขณะที่การค้นคืนสารสนเทศโดยใช้ค่าความน่าจะเป็นในการมิวเตชัน 0.10 ให้ค่าความแม่นยำปานกลางคือ 0.560 และการค้นคืนสารสนเทศโดยใช้ค่าความน่าจะเป็นในการมิวเตชัน 0.3 จะให้ค่าความแม่นยำต่ำสุดคือ 0.417 ดังแสดงในรูปที่ 2
3. ค่าความระลึก จากการทดลองค้นคืนสารสนเทศด้วยค่าความน่าจะเป็นในการครอสโอเวอร์ 0.80 และค่าความน่าจะเป็นในการมิวเตชัน 0.10 ให้ค่าความระลึกต่ำสุด คือ 0.935 การค้นคืนสารสนเทศโดยใช้ค่าความน่าจะเป็นในการมิวเตชัน 0.01 ให้ค่าความระลึกปานกลาง คือ 0.971 และการค้นคืนสารสนเทศโดยใช้ค่าความน่าจะเป็นในการมิวเตชัน 0.30 ให้ค่าความระลึกสูงสุด คือ 0.976 ดังแสดงในรูปที่ 2



1. รูปที่ 2 ค่าความแม่นยำและค่าความระลึก

4. จากตารางที่ 2 จะเห็นได้ว่าการค้นคืนโดยคิวรีบางคำทำให้ได้รับเอกสารที่ไม่เกี่ยวข้องจำนวนมาก สาเหตุเนื่องจากคำสำคัญเหล่านั้นมักปรากฏอยู่ร่วมกับคำสำคัญอื่นบางคำ ซึ่งทำให้เมื่อทำการปรับปรุงคิวรีแล้วได้รับคำสำคัญซึ่งมักจะปรากฏอยู่ด้วยกันนั้นมาด้วย แต่คำสำคัญนั้นไม่ได้บ่งชี้ถึงสาระสำคัญของเอกสารที่ผู้ใช้ต้องการโดยตรง เพียงแต่เป็นเรื่องที่มักสัมพันธ์หรือเกี่ยวข้องกัน เช่น Security มัก ปรากฏอยู่ร่วมกับหนังสือที่เกี่ยวกับ Network เมื่อทำการปรับปรุงคิวรีจนได้คิวรีที่เหมาะสมเพื่อทำการค้นคืน ปรากฏว่ามีเอกสารที่เกี่ยวกับ Network แต่ไม่เกี่ยวกับ Security โดยตรงได้รับการค้นคืนขึ้นมาด้วย วิธีการที่จะช่วยลดข้อผิดพลาดนี้ทำได้โดยใช้การป้อนกลับของผู้ใช้ (Relevance Feedback) หรือการปรับปรุงคิวรีให้เฉพาะเจาะจงมากขึ้นด้วยการใช้ตรรกะ (AND, OR, NOT) ร่วมกับคิวรีเพื่อปรับปรุงคิวรีให้สามารถค้นคืนได้ตรงตามความสนใจของผู้ใช้มากขึ้น

ตารางที่ 3 การเปรียบเทียบประสิทธิภาพ

ผลงานวิจัย	ค่าความแม่นยำ	ค่าความระลึก
วิธีการของ Kraft	0.842	0.664
วิธีการที่นำเสนอ	0.746	0.971

5. ผลการค้นคืนสารสนเทศโดยวิธีการที่นำเสนอเมื่อเปรียบเทียบกับวิธีการโดยทั่วไปพบว่า วิธีการที่นำเสนอให้ค่าความระลึกสูงกว่าวิธีการทั่วไป เนื่องจากวิธีการที่นำเสนอนอกจากจะสามารถค้นคืนเอกสารที่ค้นคืนได้ตามวิธีเปรียบเทียบแบบตรงกัน (Exact Match) แล้วระบบยังสามารถค้นคืนสารสนเทศที่เกี่ยวข้องแต่ไม่ได้ระบุโดยตรงจากคิวรีอีกด้วย และเมื่อเปรียบเทียบกับผลงานวิจัยที่ใกล้เคียงคือผลงานของ Kraft [4] ซึ่งเป็นการปรับปรุงคิวรีเพื่อใช้ในการค้นคืนสารสนเทศเช่นเดียวกันนั้น ผลปรากฏดังตารางที่ 3 คือวิธีการที่นำเสนอมีค่าความแม่นยำต่ำกว่า แต่ให้ค่าความระลึกสูงกว่าวิธีการของ Kraft และเมื่อเปรียบเทียบประสิทธิภาพ (E-Measure) ตามวิธีการของ van Rijsbergen [11] โดยรวมค่าความระลึกและค่าความแม่นยำออกมาเป็นค่าเดียว ดังสมการที่ 3

$$E = 1 - \frac{(1+b^2)PR}{b^2P + R} \qquad (3)$$

โดยที่ P คือค่าความแม่นยำ R คือค่าความระลึกและ b เป็นการวัดความสำคัญเชิงสัมพัทธ์ของค่าความระลึกและความแม่นยำของผู้ใช้ ให้ b=1 คือความสำคัญระหว่างค่าความระลึกและค่าความแม่นยำเท่ากัน ผลปรากฏว่าวิธีการที่นำเสนอมีค่า E เท่ากับ 0.156 และวิธีของ Kraft มีค่า E เท่ากับ 0.258 ซึ่งค่า E ที่ยิ่งเข้าใกล้ 0 แสดงว่ายิ่งมีประสิทธิภาพดี ดังนั้นสรุปได้ว่าวิธีการที่นำเสนอมีประสิทธิภาพดีกว่า

6. สรุป

จากผลการทดลองในเบื้องต้นจะเห็นว่าค่าความแม่นยำและค่าความระลึกจะมีลักษณะเชิงผกผันกัน การจะเลือกใช้พารามิเตอร์ใดขึ้นกับความเหมาะสมว่าต้องการใช้ค้นคืนสารสนเทศเพื่ออะไร กรณีที่ต้องการเอกสารที่มีค่าความแม่นยำสูงก็ควรจะเลือกใช้ค่าความน่าจะเป็นในการครอสโอเวอร์สูงและค่าความน่าจะเป็นในการมิวเตชันต่ำ ในขณะที่หากต้องการเอกสารที่เกี่ยวข้องมาก (ค่าความระลึกสูง) ก็อาจใช้ค่าความน่าจะเป็นในการมิวเตชันสูงและค่าความน่าจะเป็นในการครอสโอเวอร์ต่ำลง จากผลการทดลองในเบื้องต้นนี้จะเห็นว่าเราสามารถใช้จีเนติกอัลกอริทึมกับการค้นคืนสารสนเทศได้ นอกจากนี้ วิธีการที่นำเสนอเป็นการใช้คำสำคัญเป็นหลักในการค้นหา ซึ่งคำสำคัญเหล่านั้นจะเป็นตัวชี้ไปยังเอกสารอีกทีหนึ่ง ดังนั้นวิธีการที่นำเสนอจึงสามารถนำไปประยุกต์ใช้สำหรับค้นคืนสารสนเทศภาษาใด ๆ ก็ได้

สำหรับแนวทางในการวิจัยต่อไปคือทำการทดลองกับฐานข้อมูลที่มีขนาดใหญ่ขึ้น พร้อมทั้งนำเสนอเอกสารจากการค้นคืนตามลำดับของค่าความเหมาะสมที่วัดได้จากฟิตเนสฟังก์ชันซึ่งจะบ่งบอกถึงลำดับความต้องการของผู้ใช้

เอกสารอ้างอิง

[1] David, L. **Handbook of genetic algorithms.** New York : Van Nostrand Reinhold. 1991.

[2] Goldberg, D.E. **Genetic Algorithms: in Search, Optimization, and Machine Learning.** New York : Addison-Wesley Publishing Co. Inc. 1989.

[3] Korfhage, R.R. **Information Storage and Retrieval.** New York : Wiley Computer Publishing. 1997.

[4] Kraft, D.H. et. al. "The Use of Genetic Programming to Build Queries for Information Retrieval." in **Proceedings of the First IEEE Conference on Evolutional Computation**. New York : IEEE Press. PP. 468-473.

[5] Martin-Bautista, M.J. et. al. "An Approach to An Adaptive Information Retrieval Agent using Genetic Algorithms with Fuzzy Set Genes." In **Proceeding of the Sixth International Conference on Fuzzy Systems.** New York : IEEE Press. 1997. PP. 1227-1232.

[6] Salton, G. **Automatic text processing : the transformation, analysis, and retrieval of information by computer.** New York : Addison-Wesley Publishing Co. Inc. 1989.

[7] กาญจนี วงศ์วิภาพร. "การจัดตารางสอนของโรงเรียนแบบอัตโนมัติโดยจีเนติก อัลกอริทึม." วิทยานิพนธ์วิศวกรรมศาสตร์มหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2541.

[8] จุมพล พลวิชัย. "การเรียนรู้ของแขนหุ่นยนต์โดย ใช้การโปรแกรมพันธุการ." วิทยานิพนธ์ วิศวกรรมศาสตร์มหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย. 2538.

[9] ชัยวัฒน์ เจษฎาปกรณ์. "การลดทอนความเพียรพยายามเชิงคำนวณของวิธีการเรียนรู้แบบการโปรแกรมเชิงพันธุศาสตร์"

วิทยานิพนธ์วิศวกรรมศาสตร์มหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย. 2540.

[10] ธวัชชัย เอี่ยมมนัสสกุล. “การปรับปรุงประสิทธิผลของโปรแกรมหุ่นยนต์ซึ่งก่อกำเนิดโดยการโปรแกรมเชิงพันธุศาสตร์.” วิทยานิพนธ์วิศวกรรมศาสตร์มหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย. 2540.

[11] นิพนธ์ เจริญกิจการ. “การจัดเก็บและค้นคืนสารสนเทศ ฉบับปรับปรุงครั้งที่ 1.” [Online]. Available : http://web.it.kmutt.ac.th/nipon/ yllabus-temp.html. 2542.

[12] สุรเดช ตรีวิทยากรานต์. “การจัดตารางเวลาสอบหัวข้อวิจัยโดยอาศัยเทคนิคจีเนติก อัลกอริทึม.” วิทยานิพนธ์วิทยาศาสตร์มหาบัณฑิต สาขาเทคโนโลยีสารสนเทศ บัณฑิตวิทยาลัย, มหาวิทยาลัยพระจอมเกล้าธนบุรี. 2540.

Bangorn Klabbankoh received bachelor degree in education technology from King Mongkut’s University of Technology Thonburi (KMUTT) in 1997, bachelor degree in Business Administration (Marketing) from Ramkamhaeng University in 1999 and master degree in information technology from King Mongkut’s Institute of Technology Ladkrabang (KMITL) in 2000. Her research interests are Genetic Algorithm, Information Retrieval and Expert System.

ผู้ช่วยศาสตราจารย์ เอื้อน ปิ่นเงิน สำเร็จการศึกษาระดับปริญญาตรีจากมหาวิทยาลัยศรีนครินทรวิโรฒสาขาคณิตศาสตร์ ปริญญาโทจากจุฬาลงกรณ์มหาวิทยาลัย และ Qregon State University สาขาคอมพิวเตอร์ และได้สำเร็จปริญญาเอกทางคอมพิวเตอร์จาก University of Nebraska ประเทศสหรัฐอเมริกา ปัจจุบันเป็นอาจารย์ประจำภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ความก้าวหน้าของการพัฒนาระบบระบุผู้พูดภาษาไทย

Thai Language Speaker Identification System: Development Progress[1]

ชัย วุฒิวิวัฒน์ชัย, สุทัศน์ แซ่ตั้ง และวารินทร์ อัจฉริยะกุลพร
คณะนักวิจัยและพัฒนาระบบระบุผู้พูดสำหรับภาษาไทย[2]
หน่วยปฏิบัติการวิจัยและพัฒนาวิศวกรรมภาษาและซอฟต์แวร์
ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ
สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ
539/2 อาคารมหานครยิบซั่ม ชั้น 22 ถนนศรีอยุธยา แขวงพญาไท เขตราชเทวี กรุงเทพฯ 10400

ABSTRACT -- Speaker identification for Thai language project has been initiated by the National Electronics and Computer Technology Center (NECTEC) since 1999. The first objective is to research and develop a text-dependent closed-set speaker identification system in the office environment. The speaking texts for this system are isolated digit utterances 0-9 and their concatenation. This paper gives an overview of the system, explains the route of the past 1-year research history, and some details of the latest identification system, which achieves the best performance of 92.30% for isolated digit "0" and enhances to 98% for 3-concatenated digit.

KEY WORDS -- Speaker Identification, Text Dependent, Closed Set, Thai Language

บทคัดย่อ -- โครงการระบบระบุผู้พูดสำหรับภาษาไทย (Speaker Identification for Thai Language) ของศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ ได้ริเริ่มขึ้นในปีงบประมาณ 2542 โดยเบื้องต้นมุ่งเน้นการวิจัยและพัฒนาระบบระบุผู้พูดที่ใช้กับภาษาไทยแบบกำหนดคำพูดตายตัว (Text dependent) เป็นระบบปิด (Closed set system) และใช้ในสภาพแวดล้อมสำนักงาน (Office environment) คำพูดที่ใช้ในการวิจัยเป็นเสียงตัวเลขโดด 0-9 และตัวเลขโดดต่อกัน บทความฉบับนี้เป็นการสรุปผลการวิจัยโดยนำเสนอภาพรวมของระบบระบุผู้พูด แสดงรายละเอียดของผลงานวิจัยในช่วง 1 ปีที่ผ่านมา รวมทั้งนำเสนอผลงานความก้าวหน้าล่าสุด พร้อมทั้งรายละเอียดของระบบระบุผู้พูดที่ใช้กับผู้พูดจำนวน 50 คน ซึ่งได้ผลอัตราการระบุผู้พูดสูงที่สุด 92.30% เมื่อใช้เสียงตัวเลข 0 และเพิ่มขึ้นเป็น 98% เมื่อใช้เสียงตัวเลขโดดต่อกัน 3 ตัว

คำสำคัญ -- การระบุผู้พูด, กำหนดคำพูดตายตัว, ระบบปิด, ภาษาไทย

---

[1] บทความนี้ตีพิมพ์ครั้งแรกในเอกสารประกอบการประชุมวิชาการของศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ ปี 2543 หน้า 496-510 และได้รับรางวัลบทความวิชาการดีเด่น

[2] คณะนักวิจัยประกอบด้วย ดร.จุฬารัตน์ ตันประเสริฐ หัวหน้าโครงการ, นายวศิน สินธุภิญโญ, นายเปรมนาถ ดูเบ, นายสุทัศน์ แซ่ตั้ง, นายวารินทร์ อัจฉริยะกุลพร, นายชัย วุฒิวิวัฒน์ชัย และนายศวิต กาสุริยะ

1. บทนำ

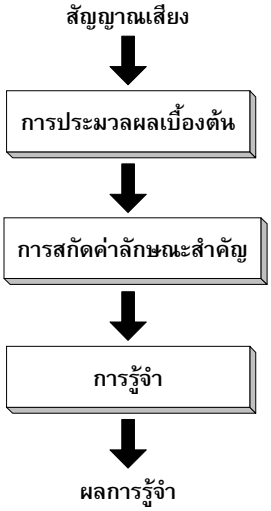
ในปัจจุบันระบบที่ใช้องค์ประกอบและลักษณะของบุคคลมาระบุตัวบุคคลนั้นๆ (Biometrics personal identification system) เพื่อใช้ในระบบรักษาความปลอดภัย แทนการป้อนรหัสผ่านทางแป้นพิมพ์ (Password) หรือการใช้บัตรแถบแม่เหล็ก (Magnetic card) เป็นที่นิยมมาก เช่น การตรวจสอบลายนิ้วมือ (Fingerprints) การตรวจสอบรูปแบบม่านตา (Retinal patterns) หรือจะเป็นการตรวจสอบใบหน้า (Face recognition) เป็นต้น เหตุผลประการหนึ่งที่ระบบดังกล่าวได้รับความนิยมเพราะยากต่อการปลอมแปลง ในขณะที่การใช้รหัสผ่าน หรือบัตรแถบแม่เหล็กนั้นง่ายต่อการถูกลักขโมย รวมทั้งอาจจะลืมรหัสผ่าน หรือลืมนำบัตรติดตัวมาด้วย

ระบบการรู้จำผู้พูด (Speaker recognition system) ก็เป็นหนึ่งในเทคโนโลยีดังกล่าว ที่ได้รับความสนใจนำมาใช้ในการระบุตัวบุคคล [1,2] นอกเหนือจากระบบระบุตัวบุคคลอื่นๆ การรู้จำผู้พูดสามารถแบ่งออกได้เป็น 2 ประเภทหลักๆ คือ การรับรองผู้พูด (Speaker verification) ซึ่งเป็นการตรวจสอบผู้พูดว่าเป็นบุคคลเดียวกับบุคคลที่กำหนดหรือไม่ และการระบุผู้พูด (Speaker identification) ซึ่งจะทำการตรวจสอบผู้พูดว่าเป็นใคร [3] นอกจากนี้การระบุผู้พูดยังแบ่งได้เป็น 2 อย่างคือ การระบุผู้พูดแบบปิด (Closed-set) เป็นการระบุว่าผู้พูดเป็นบุคคลใดในกลุ่มบุคคลที่กำหนด ในขณะที่การระบุผู้พูดแบบเปิด (Open-set) เป็นการระบุว่าผู้พูดเป็นบุคคลใดในกลุ่มบุคคลที่กำหนด หรือเป็นบุคคลนอกกลุ่ม ระบบการรู้จำผู้พูดยังสามารถแบ่งได้ตามข้อความที่พูดคือ แบบกำหนดคำ หรือประโยคให้พูด (Text dependent) และแบบไม่กำหนดคำ หรือประโยคให้พูด (Text independent) หรือแบ่งตามสถานที่ใช้งาน คือในสภาพแวดล้อมของสำนักงาน (Office environment) และสภาพแวดล้อมทางโทรศัพท์ (Telephone environment)

สำหรับงานวิจัยนี้ เป็นการวิจัยและพัฒนาระบบระบุผู้พูดสำหรับภาษาไทยแบบกำหนดคำพูดตายตัว และเป็นระบบปิด ใช้ในสภาพแวดล้อมสำนักงาน ถือได้ว่าเป็นงานวิจัยในยุคเริ่มแรกของการวิจัยระบบรู้จำผู้พูดสำหรับภาษาไทย และบทความฉบับนี้เป็นบทสรุปความก้าวหน้าของงานวิจัยใน 1 ปีที่ผ่านมา โดยแบ่งหัวข้อดังนี้ หัวข้อที่ 2 จะกล่าวถึงภาพรวมวิธีการระบุผู้พูด โดยให้รายละเอียดคร่าวๆ ของส่วนประกอบต่างๆ ในระบบ หัวข้อที่ 3 จะกล่าวถึงขั้นตอนของงานวิจัยที่ผ่านมาโดยแบ่งแยกเป็นงานวิจัยในส่วนต่างๆ ของระบบ หัวข้อที่ 4 จะกล่าวถึงผลงานล่าสุดที่ให้ผลการระบุผู้พูดสูงที่สุด หัวข้อที่ 5 จะสรุปปัญหา ข้อเสนอแนะและงานวิจัยที่จะดำเนินการต่อไปในอนาคต และสรุปเนื้อหาของบทความนี้ในหัวข้อที่ 6

2. ภาพรวมของระบบระบุผู้พูด

หลักการโดยทั่วไป สำหรับการสร้างระบบระบุผู้พูดได้แสดงไว้ในรูปที่ 1 [3] ประกอบด้วยการประมวลผลเบื้องต้น (Preprocessing) การสกัดค่าลักษณะสำคัญ (Feature extraction) และการรู้จำ (Recognition)



รูปที่ 1. หลักการโดยทั่วไปของระบบระบุผู้พูด

2.1 การประมวลผลเบื้องต้น (Preprocessing)

สัญญาณเสียงที่ผ่านการแปลงสัญญาณเป็นดิจิตอลแล้ว จะนำมาผ่านขั้นตอนการประมวลผลเบื้องต้น ซึ่งประกอบด้วยขั้นตอนต่างๆ ดังนี้

1. การกรองทางความถี่ (Filtering) เป็นขั้นตอนในการกรองสัญญาณในช่วงความถี่ที่ไม่ต้องการออกโดยอาศัยตัวกรองแบบดิจิตอล
2. การตัดหัว-ท้ายเสียง (Endpoint detection) เป็นขั้นตอนในการกำหนดจุดเริ่มต้นและจุดสิ้นสุดของเสียง โดยการแยกส่วนที่เป็นคำพูดออกจากส่วนที่ไม่ใช่คำพูด วิธีในการตัดหัว-ท้ายเสียงมีหลายวิธี เช่น ใช้ค่าระดับพลังงาน (Energy level) ใช้อัตราการตัดศูนย์ (Zero-crossing rate) เป็นต้น
3. การนอร์มอลไลซ์ทางเวลา (Time normalization) เป็นขั้นตอนการเพิ่ม หรือลดขนาดความยาวของสัญญาณในเชิงเวลา เพื่อปรับแต่งขนาดความยาวของสัญญาณให้เหมาะสมตามต้องการ ทั้งนี้จะขึ้นอยู่กับกระบวนการในการรู้จำว่าจำเป็นต้องทำการนอร์มอลไลซ์สัญญาณให้เท่ากันหรือไม่ วิธีในการนอร์มอลไลซ์ทางเวลามีหลายวิธี เช่น การเปลี่ยนอัตราการชักตัวอย่าง (Sampling rate changing) การประมาณค่าในช่วงเชิงเส้น (Linear interpolation) [4] และการเหลื่อมและรวมส่วนย่อยแบบซิงโครไนซ์ (Synchronized overlap-and-add) [5] เป็นต้น

2.2 การสกัดค่าลักษณะสำคัญ (Feature)

การสกัดค่าลักษณะสำคัญ คือการวิเคราะห์หาค่าที่จะใช้แทนสัญญาณเสียง เพื่อนำไปใช้ในขั้นตอนการรู้จำ แบ่งได้เป็น 3 กลุ่มหลัก กลุ่มแรก

เป็นค่าลักษณะสำคัญระดับสูง (High level feature) ได้แก่ สำเนียงการพูด รูปแบบในการพูด และความเร็วในการพูด เป็นต้น ในกลุ่มที่สอง จะใช้ค่าลักษณะสำคัญทางฉันทลักษณ์ (Prosodic feature) เช่น ค่าความถี่มูลฐาน (Fundamental frequency) ความถี่ฟอร์แมนท์ (Formant frequency) และระดับพลังงาน (Energy profile) เป็นต้น ถึงแม้ว่าค่าลักษณะสำคัญแบบนี้จะมีประสิทธิภาพสูงในการรู้จำ แต่ยากในการสกัดจากสัญญาณ กลุ่มสุดท้ายเรียกว่าค่าลักษณะสำคัญแบบเอนเวโลปของสเปกตรัม (Spectral envelop feature) [6] เป็นกลุ่มที่นิยมใช้กันมาก เนื่องจากค่าลักษณะสำคัญส่วนใหญ่สำหรับการรู้จำเสียงจะรวมอยู่ในข้อมูลเชิงสเปกตรัมนี้ อีกทั้งยังง่ายและสะดวกในการคำนวณหาค่าด้วย ตัวอย่างค่าลักษณะสำคัญแบบนี้ได้แก่ สัมประสิทธิ์การประมาณพันธะเชิงเส้น (Linear prediction coefficients: LPC), สัมประสิทธิ์เซปสตรัม (Cepstral coefficient) และพัฒนาการอีกมากมายจากเซปสตรัมปกติ [7] อาทิเช่น สัมประสิทธิ์เซปสตรัมบนสเกลเมล (Mel frequency cepatral coefficients: MFCC) เซปสตรัมแบบหักลบค่าเฉลี่ย (Cepstral mean substraction: CMS) และเซปสตรัมแบบผ่านตัวกรองภายหลัง (Post filtered cepsturm: PFL) เป็นต้น นอกจากนั้น ยังมีการคำนวณค่าการเปลี่ยนแปลง (Derivative หรือ Delta) ของสัมประสิทธิ์เหล่านี้มาใช้เป็นค่าลักษณะสำคัญเพิ่มเติมได้ด้วย

สำหรับการคำนวณค่าลักษณะสำคัญแบบเอนเวโลปของสเปกตรัมจะมีขั้นตอนดังนี้ [3]

1. การเน้นสัญญาณขั้นต้น (Preemphasis) เป็นขั้นตอนในการบีบอัดสัญญาณเสียงโดยนำสัญญาณเสียงผ่านตัวกรองลำดับหนึ่ง (First-order filter) ซึ่งจะเพิ่มอัตราส่วนสัญญาณต่อสัญญาณรบกวน (Signal to noise ratio)
2. การแบ่งเป็นส่วนย่อย (Frame) เป็นขั้นตอนในการแบ่งสัญญาณเสียงเป็นส่วนย่อย ขนาดความยาวประมาณ 10 – 40 มิลลิวินาที ซึ่งทำให้สัญญาณเสียงมีคุณสมบัติเปลี่ยนแปลงตามเวลาน้อยมาก หรือไม่มีเลย เพื่อให้สามารถสร้างแบบจำลองการกระจายของหน่วยสัญญาณเสียงย่อยทางสถิติได้
3. การลดขอบด้วยฟังก์ชันหน้าต่างสำหรับปรับสัญญาณให้ราบเรียบ (Smoothing window)
4. การสกัดค่าลักษณะสำคัญ (Feature extraction) ในส่วนนี้ จะทำการคำนวณค่าลักษณะสำคัญของสัญญาณเสียงในแต่ละส่วนย่อย ผลลัพธ์อยู่ในรูปแบบของเวกเตอร์ของค่าลักษณะสำคัญ (Feature vector) สำหรับแต่ละส่วนย่อย

2.3 การรู้จำ (Recognition)

ขั้นตอนนี้ประกอบด้วย 2 หน้าที่หลัก คือการนำเวกเตอร์ของค่าลักษณะสำคัญของสัญญาณเสียง ที่อยู่ในชุดอ้างอิงหรือชุดฝึกฝน มาทำการเรียนรู้ เมื่อเรียนรู้แล้วเวกเตอร์ของสัญญาณเสียงที่ต้องการทดสอบการรู้จำ จะถูกนำเข้ามาเทียบเคียงเพื่อรู้จำ ขั้นตอนในการเรียนรู้นั้นขึ้นอยู่กับวิธีในการรู้จำของระบบนั้นๆ บางวิธีก็เพียงแค่เก็บข้อมูลชุดเรียนรู้ไว้เปรียบเทียบกับข้อมูลชุดทอดสอบเท่านั้น เช่น วิธีการรู้จำแบบหาค่าระยะห่างยูคลิเดียน (Euclidean distance) วิธีไดนามิกไทม์วาร์ปปิง (Dynamic time warping: DTW) [6] เป็นต้น ในขณะที่บางวิธี จะนำข้อมูลชุดเรียนรู้ไปแปลงเป็นค่าอ้างอิงที่ต้องการ เช่น โครงข่ายประสาทเทียม (Artificial neural networks: ANN) [8] จะนำข้อมูลชุดเรียนรู้ไปผ่านโครงข่ายที่สร้างขึ้นเพื่อจดจำรูปแบบ และเก็บเป็นค่าน้ำหนัก (Weight) แทน วิธีควอนไตซ์แบบเวกเตอร์ (Vector quantization: VQ) [9] ซึ่งจะแทนเวกเตอร์ทั้งหมดของแต่ละสัญญาณเสียงอ้างอิงด้วยเวกเตอร์จำนวนไม่มาก หรือการใช้แบบจำลองฮิดเดนมาร์คอฟ (Hidden markov model: HMM) [6,9] โดยนำข้อมูลชุดฝึกฝน ไปผ่านแบบจำลองที่สร้างขึ้นเพื่อจดจำรูปแบบ และเก็บค่าทางสถิติและค่าความน่าจะเป็นของแต่ละสถานะไว้ เป็นต้น แต่ทั้งหมดจะมีพื้นฐานอยู่ที่การคำนวณระยะห่างของรูปแบบที่จะรู้จำ และนำค่าระยะห่างที่ได้ไปใช้รู้จำตามแต่ละวิธีนั้นๆ

การเลือกใช้วิธีการรู้จำ ขึ้นอยู่กับข้อกำหนดของงาน เช่น วิธี DTW และ ANN เหมาะสมกับระบบแบบกำหนดคำพูดตายตัว ในขณะที่วิธี VQ และ HMM จะเหมะสมกับระบบงานที่เป็นแบบไม่กำหนดคำพูดมากกว่า [1,9]

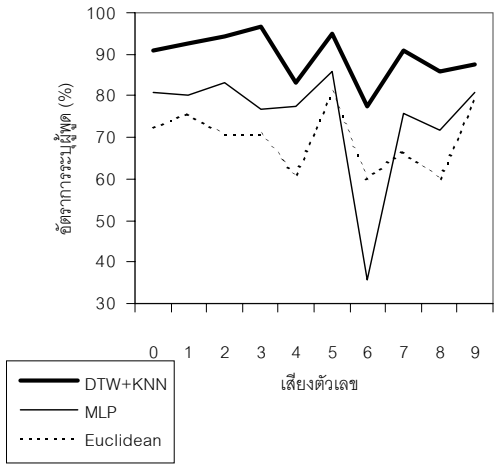
3. เส้นทางการวิจัยที่ผ่านมา

เนื่องจากงานด้านการระบุผู้พูดสำหรับภาษาไทยในประเทศไทยยังไม่มีผลงานวิจัยมากนัก คณะวิจัยจึงเริ่มศึกษาแนวทางจากบทความการระบุผู้พูดจากภาษาต่างประเทศเพื่อเป็นแนวทางในการดำเนินการ งานวิจัยเริ่มแรกควรใช้ระบบระบุผู้พูดแบบกำหนดคำพูดตายตัวเพื่อให้ง่ายต่อการวิจัยและไม่ซับซ้อนเกินไปนัก โดยทดสอบกับเสียงของตัวเลขโดด 0 ถึง 9 ของภาษาไทย ซึ่งคาดว่าจะสามารถประยุกต์นำไปใช้กับระบบรักษาความปลอดภัย ระบุผู้พูด หรือพิสูจน์ผู้พูดได้ เช่น การระบุผู้พูดจากรหัสประจำตัว เป็นต้น

3.1 การทดลองขั้นต้น

ในขั้นเริ่มต้นของงานวิจัย ได้มีความพยายามในการเลือกระบบรู้จำที่จะนำมาใช้ โดยเปรียบเทียบระบบรู้จำอย่างน้อย 3 ระบบ คือวิธีหาระยะห่างแบบยูคลิเดียน, วิธี DTW โดยใช้วิธีตัดสินใจแบบพิจารณาจุดใกล้ K จุด (K-nearest neighbor: K-NN) [2], และการใช้ ANN ชนิดเพอเซปตรอนหลายชั้น (Multilayer perceptron network: MLP) ร่วมกับการเรียนรู้แบบแพร่กระจายย้อนกลับ (Back-propagation) [8]

การทดลองทำการระบุผู้พูดจำนวน 20 คน (ชาย 11 คน หญิง 9 คน) โดยผู้พูดแต่ละคนจะต้องอัดเสียงพูดตัวเลขโดด 0-9 จำนวน 10 ครั้งต่อสัปดาห์ เป็นเวลา 5 สัปดาห์ แบ่งเสียงจาก 3 สัปดาห์แรกเป็นชุดฝึกฝน ที่เหลือเป็นชุดทดสอบ สัญญาณเสียงจะถูกนำมาผ่านผ่านการนอร์มอลไลซ์ทางเวลาเฉพาะในกรณีของระบบรู้จำแบบยูคลิเดียนและ ANN หลังจากนั้น แบ่งเป็นส่วนย่อยๆ ละ 20 มิลลิวินาที เหลื่อมส่วนย่อยละ 5 มิลลิวินาที และใช้ค่าลักษณะสำคัญแบบ LPC ขนาด 10 อันดับ ผลการทดลอง [10] แสดงดังรูปที่ 2



รูปที่ 2. กราฟเปรียบเทียบผลอัตราการระบุผู้พูดสำหรับวิธี DTW, ANN และระยะห่างแบบยูคลิเดียน

ผลการทดลองชี้ให้เห็นถึงความสามารถของ DTW ซึ่งให้ผลการระบุผู้พูดสูงสุดถึง 96.67% กับเสียงพูดเลข 5 และผลอัตราการระบุผู้พูดเฉลี่ยสำหรับทุกเสียงตัวเลขเท่ากับ 89.42% ในขณะที่ ANN ให้อัตราการระบุผู้พูดสูงสุด 85.83% กับเสียงพูดเลข 3 และอัตราการระบุผู้พูดเฉลี่ย 74.83% ส่วนการใช้ระยะห่างแบบยูคลิเดียนให้ผลต่ำที่สุด โดยมีอัตราการระบุผู้พูดเฉลี่ยเพียง 69.75% เท่านั้น

จากผลการทดลอง ส่วนหนึ่งที่วิเคราะห์ได้คือการใช้วิธีนอร์มอลไลซ์ทางเวลา น่าจะเป็นวิธีที่ลดประสิทธิภาพของการรู้จำได้มาก อย่างไรก็ตาม DTW ได้กลายมาเป็นวิธีที่ถูกนำมาวิจัยและพัฒนาต่อ รวมทั้งการวิจัยวิธีการอื่นๆ ของระบบรู้จำที่หลีกเลี่ยงการนอร์มอลไลซ์ทางเวลา ในขณะเดียวกันการทดลองเพื่อเลือกค่าลักษณะสำคัญที่ให้ประสิทธิภาพสูงขึ้นก็จะทำควบคู่กันไปด้วยกับวิธีการรู้จำแบบต่างๆ เนื่องจากเราไม่สามารถบอกได้ว่า จะใช้ค่าลักษณะสำคัญแบบใดจึงเหมาะสมกับเทคนิคในการเทียบเคียงรูปแบบสัญญาณแต่ละแบบ

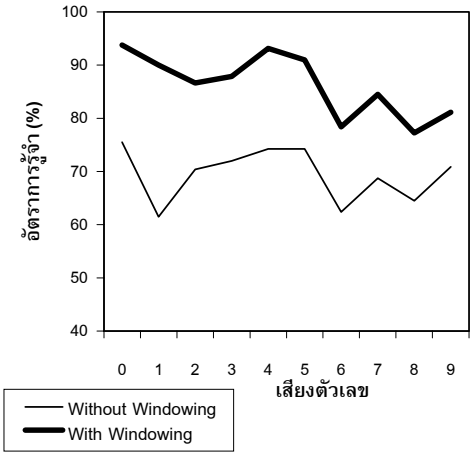
3.2 การวิจัยเกี่ยวกับระบบรู้จำ

หลังจากการทดลองขั้นต้นเกี่ยวกับระบบรู้จำที่จะนำมาใช้ดังได้กล่าวมาแล้ว งานวิจัยก็หันมามุ่งเน้นที่การพัฒนาระบบรู้จำโดยได้ศึกษารายละเอียดของวิธีการรู้จำนั้นๆ เพื่อปรับเปลี่ยนค่าพารามิเตอร์ และขั้นตอนกระบวนการต่างๆ เพื่อให้เหมาะสมกับสัญญาณเสียงภาษาไทย เพื่อเพิ่มผลอัตราการรู้จำให้สูงขึ้น และสามารถรองรับจำนวนผู้พูดได้มากขึ้นโดยไม่มีผลกระทบต่ออัตราการรู้จำ

<u>การทดลองเกี่ยวกับ ANN</u>

ปัญหาสำคัญประการหนึ่งของ ANN แบบ MLP ซึ่งอาศัยการเรียนรู้แบบแพร่กระจายย้อนกลับ คือจะต้องกำหนดจำนวนข้อมูลในชั้นข้อมูลเข้าให้เท่ากันทุกๆ รูปแบบที่จะรู้จำ ซึ่งจำเป็นต้องใช้การนอร์มอลไลซ์ทางเวลามาช่วย ทำให้สูญเสียลักษณะสำคัญบางประเภทที่จำเป็นสำหรับการรู้จำไป อัตราการรู้จำจึงไม่สูงเท่าที่ควร เพื่อไม่ให้ลักษณะสำคัญสูญเสียระหว่างการทำกระบวนการนอร์มอลไลซ์ทางเวลา และเพื่อประหยัดเวลาในการรู้จำ คณะนักวิจัยฯ จึงได้คิดหาวิธีในการส่งข้อมูลลักษณะสำคัญเข้าสู่ชั้นข้อมูลเข้าเพื่อใช้ในการฝึกหัดใหม่ ซึ่งทำให้ไม่มีความจำเป็นในการทำนอร์มอลไลซ์ทางเวลาอีกต่อไป โดยเปลี่ยนให้มีการส่งข้อมูลเข้าแบบหน้าต่าง (Windowing technique) [11] รายละเอียดของกระบวนการจะกล่าวในหัวข้อต่อไป

การทดลองสำหรับระบุผู้พูดจำนวน 20 คนเช่นเดิม โดยใช้ค่าลักษณะสำคัญแบบเซปสตรัมที่คำนวณมาจาก LPC (Linear predictive coding derived cepstrum, LPCC) ขนาด 15 อันดับ ผลการทดสอบแสดงดังรูปที่ 3 ยืนยันว่า วิธีการส่งข้อมูลเข้าแบบหน้าต่างนี้ให้ผลการรู้จำที่สูงกว่าแบบเดิมโดยเฉลี่ยถึง 15% [8] กล่าวคือให้อัตราระบุผู้พูดได้สูงสุดถึง 93.75% สำหรับเสียงตัวเลข 0 และให้อัตราการรู้จำเฉลี่ยสูงถึง 86.37%



รูปที่ 3. กราฟเปรียบเทียบผลอัตราการระบุผู้พูดด้วยวิธี ANN ปกติและแบบใช้เทคนิค Windowing

<u>การทดลองเกี่ยวกับ DTW</u>

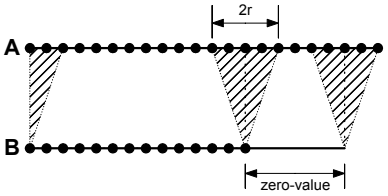
วิธีการรู้จำแบบ DTW เป็นอีกวิธีหนึ่งที่ได้ศึกษาอย่างละเอียด หลังจากได้ศึกษาวิธีการรู้จำแบบ DTW แล้วพบว่าพารามิเตอร์ต่างๆ ใน DTW ได้แก่

การเลือกค่า r หรือกรอบของจุดที่อนุญาตให้มีการจับคู่จุดได้ (Time alignment window) และจำนวนจุดของการก้าวแต่ละครั้งให้เหมาะสม การกำหนดค่า r ไว้คงที่ จะส่งผลกระทบกับการจับคู่ลำดับ 2 ลำดับที่มีความยาวต่างกันมากๆ นอกจากนี้ถ้ามีจำนวนชุดอ้างอิงมากก็จะสูญเสียเวลาในการคำนวณระยะห่างมาก และถ้ามีจำนวนผู้พูดมากก็จะเสียเวลาในการคำนวณมากเช่นกัน

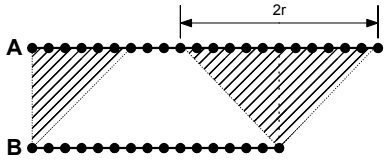
จากปัญหาต่างๆ ดังกล่าว จึงได้มีการทดลอง 3 ส่วน ส่วนแรกได้ทำการทดลองเพื่อปรับค่า r ที่เหมาะสม [12] โดยทำการทดลองระบุผู้พูดจำนวน 20 คน ใช้ค่าลักษณะสำคัญแบบ LPC ขนาด 10 อันดับ ปรากฏว่าค่า r ที่เหมาะสมขึ้นอยู่กับความยาวของเสียงที่ใช้ในการระบุผู้พูด สำหรับเสียงตัวเลขโดด ค่า r เท่ากับ 5 ให้ผลการระบุได้สูง และค่า r ควรจะเพิ่มขึ้นเมื่อใช้เสียงพูดยาวขึ้น

ส่วนที่ 2 เป็นการเสนอขั้นตอนกระบวนการในการกำหนดค่า r รวมทั้งแก้ไขปัญหาความแตกต่างของความยาวของลำดับที่เทียบเคียงกัน ในส่วนนี้ได้เสนอเทคนิคการเทียบเคียงด้วย DTW 3 เทคนิค ดังแสดงในรูปที่ 4 เทคนิคแรกเป็นการกำหนดค่า r คงที่และมีการเพิ่มค่าศูนย์ต่อท้ายสำหรับลำดับที่สั้นกว่า เทคนิคที่ 2 เป็นการกำหนดค่า r ให้เท่ากับความแตกต่างของความยาวของลำดับที่เทียบเคียงกัน เทคนิคสุดท้ายเป็นการยืดจุดที่ควรจะเทียบเคียงกัน ตามสัดส่วนของความยาวแล้วจึงกำหนดค่า r ให้คงที่ค่าหนึ่ง ผลการทดลอง [13] ปรากฏว่าในการระบุผู้พูดจำนวน 50 คน โดยใช้ค่าลักษณะสำคัญแบบ LPCC ขนาด 15 อันดับ เทคนิคที่ 1 และ 3 ให้ผลการระบุผู้พูดเฉลี่ยได้ใกล้เคียงกันกล่าวคือ 84.53% และ 84.29% ตามลำดับ เทคนิคที่ 1 ให้ผลดีกว่าเล็กน้อยและยังใช้จำนวนการคำนวณน้อยกว่าอีกด้วย

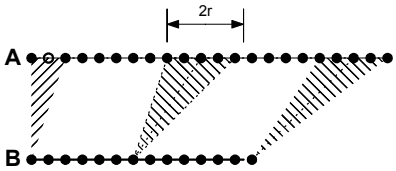
ส่วนสุดท้ายเป็นการทดลองปรับเปลี่ยนจำนวนเสียงอ้างอิงจากเดิมใช้เสียงจาก 3 สัปดาห์แรก (30 เสียง) เป็น 20 และ 10 เสียงโดยคัดแบบคละกันจาก 3 สัปดาห์แรก ผลการทดลอง [13] ปรากฏให้เห็นว่าสำหรับการระบุผู้พูด 50 คน และใช้ค่าลักษณะสำคัญแบบ LPCC ขนาด 15 อันดับ จำนวนเสียงอ้างอิงเท่ากับ 20 ให้ผลการทดลองได้ถึง 84.61% ซึ่งสูงกว่าค่าอื่น แสดงให้เห็นว่าการใช้ชุดอ้างอิงจำนวนมากอาจทำให้ระบบเกิดความสับสนได้มากขึ้น



ก. เทคนิคที่ 1



ข. เทคนิคที่ 2



ค. เทคนิคที่ 3

รูปที่ 4. ขั้นตอนกระบวนการ DTW 3 เทคนิค

<u>การทดลองระบบรู้จำแบบอื่นๆ</u>

นอกเหนือจาก ANN และ DTW แล้ว ยังได้มีความพยายามใช้ระบบรู้จำแบบอื่น ได้แก่ วิธี VQ, วิธี HMM แบบไม่ต่อเนื่อง (Discrete hidden markov model: DHMM) และวิธีแบบจำลองส่วนผสมแบบเกาส์ (Gaussian mixture model: GMM) ซึ่งกำลังอยู่ในขั้นตอนการทดลอง

<u>การทดลองวิธีการตัดสินใจเมื่อใช้เสียงตัวเลขต่อกัน</u>

การทดลองอีกส่วนหนึ่ง คือการเพิ่มความยาวของคำพูดที่ใช้ในระบบระบุผู้พูดโดยใช้เสียงตัวเลขโดดต่อกัน เพื่อเพิ่มประสิทธิภาพของการระบุผู้พูด โดยคัดเลือกเอาค่าลักษณะสำคัญของเสียงตัวเลขโดดที่ให้อัตราการระบุผู้พูดสูงที่สุดอันดับแรกๆ มาต่อกันแล้วจึงเข้าระบบรู้จำวิธีการนี้จะให้ผลอัตราการรู้จำสูงขึ้นกว่าการใช้เสียงตัวเลขโดด อย่างไรก็ตามการใช้ตัวเลขต่อกันนี้ จะทำให้ระบบใช้เวลาในการประมวลผลมากขึ้นโดยเฉพาะอย่างยิ่ง ระบบที่ใช้ DTW ซึ่งต้องกำหนดค่า r ให้กว้างขึ้น ยิ่งเป็นการเพิ่มเวลาในการรู้จำมากขึ้นไปอีก วิธีการหนึ่งสำหรับระบบระบุผู้พูดที่ใช้ DTW และ K-NN ซึ่งได้เสนอไว้ใน [13] คือการรวมเสียงอ้างอิง K เสียงที่ได้จากการระบุผู้พูดด้วยตัวเลขโดดแต่ละตัว เช่น ถ้าสำหรับตัวเลขโดด ใช้กฎการตัดสินใจแบบ 5-NN เมื่อใช้ตัวเลข 3 ตัวต่อกันในระบบระบุผู้พูด จะทำการคำนวณ 5-NN ของตัวเลขโดดแต่ละตัวแล้วจึงนำมารวมกันพิจารณาแบบ 15-NN เป็นต้น

การทดลองระบุผู้พูดจำนวน 50 คนโดยใช้ LPCC ขนาด 15 อันดับ กับตัวเลขต่อกัน 3 ตัวด้วย DTW ผลปรากฏว่าวิธีตัดสินใจแบบใหม่ให้ผลอัตราการระบุผู้พูด 96.10% ในขณะที่วิธีแบบเก่า คือการนำค่าลักษณะสำคัญของแต่ละเลขมาต่อกันก่อน โดยใช้ r = 20 จะให้อัตราการรู้จำ 95.40% ซ้ำยังใช้เวลาในการประมวลผลนานกว่ามาก หลักการเดียวกันนี้สามารถนำไปใช้ได้ในระบบระบุผู้พูดที่ใช้ ANN แบบใหม่ได้เช่นกัน

3.3 การวิจัยเกี่ยวกับค่าลักษณะสำคัญของเสียง

หลังจากได้ทดลองเกี่ยวกับระบบที่ใช้ในการเทียบเคียงพอสมควรแล้ว คณะนักวิจัยได้เริ่มหันมาพิจารณาค่าลักษณะสำคัญที่เหมาะสมสำหรับระบบแต่ละระบบ ค่าลักษณะสำคัญที่เป็นที่นิยม ได้รับการพิสูจน์แล้วว่ามีประสิทธิภาพสูง สำหรับการรู้จำผู้พูดหรือรู้จำเสียงพูดได้ถูกนำมาใช้ในการทดลองเปรียบเทียบ

ในช่วงต้นของงานวิจัย ได้มีการใช้ค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น (Linear prediction coefficient: LPC) เป็นค่าลักษณะสำคัญ ต่อมาจึงหันมาใช้ค่าสัมประสิทธิ์เซปสตรัม (Cepstral coefficient) ชนิดที่คำนวณมาจาก LPC เรียกว่า LPCC (Linear predictive coding derived cepstrum) โดยมีการทดลองที่แสดงไว้ใน [14] เป็นการทดลองระบุผู้พูดจำนวน 20 คน โดยใช้ค่าลักษณะสำคัญแบบ LPC และ LPCC ขนาด 10 และ 15 อันดับ ใช้ระบบรู้จำแบบ DTW และ K-NN ผลการทดลองดังตารางที่ 1 แสดงให้เห็นอย่างชัดเจนว่า LPCC ให้ผลการระบุผู้พูดสูงกว่า LPC มากทั้งแบบ 10 และ 15 อันดับ สำหรับ LPCC แล้ว ใช้ขนาด 15 อันดับให้ผลดีกว่า 10 ลำดับโดยให้ผลการระบุผู้พูดเฉลี่ยถึง 86.28%

ตารางที่ 1. การทดลองเปรียบเทียบอัตราการรู้จำของระบบที่ใช้ LPC และ LPCC ขนาด 10 และ 15 อันดับ

ตัวเลข	อัตราการระบุผู้พูด (%)			
	LPC		LPCC	
	10	15	10	15
0	84.00	83.75	90.50	91.00
1	71.00	68.75	86.00	89.75
2	64.75	69.50	83.00	87.00
3	74.25	67.75	83.75	85.00
4	77.50	68.50	91.00	93.00
5	70.50	70.75	90.50	91.50
6	69.00	65.25	82.25	85.75
7	60.75	56.75	81.75	84.75
8	48.50	49.50	64.25	70.00
9	70.50	66.50	86.25	85.00
เฉลี่ย	69.08	66.70	83.93	86.28

ในงานวิจัยถัดมา ค่าลักษณะสำคัญอีกหลายชนิดโดยเฉพาะค่าลักษณะสำคัญที่อยู่ในกลุ่มของเซปสตรัมได้ถูกนำขึ้นมาทดลอง อาทิเช่น เซปสตรัมแบบหักลบค่าเฉลี่ย (Cepstral mean substraction: CMS) เซปสตรัมแบบให้น้ำหนักที่ปรับส่วนประกอบได้ (Adaptive component weighted cepstrum: ACW) เซปสตรัมบนสเกลเมล (Mel frequency cepstral coefficient: MFCC) และเซปสตรัมแบบผ่านตัวกรองภายหลัง (Post filtered cepstrum: PFL) ในจำนวนนี้ ค่าลักษณะสำคัญที่ให้ผลการระบุผู้พูดได้สูงได้แก่ MFCC และ PFL จึงมีการทดลองเปรียบเทียบค่าลักษณะสำคัญ 3 ค่าคือ LPCC, MFCC และ PFL กับการระบุผู้พูดจำนวน 50 คน และกำหนดอันดับของค่าลักษณะสำคัญให้คงที่ที่ 15 อันดับ ปรากฏว่าทั้งการทดลองโดยใช้ ANN ที่ป้อนข้อมูลแบบหน้าต่าง และการทดลองโดยใช้ DTW และ K-NN [14] ผลการระบุผู้พูดเฉลี่ยจะสูงที่สุดเมื่อใช้ MFCC ซึ่งสูงกว่า PFL เพียงเล็กน้อยในขณะที่ LPCC ให้ผลต่ำที่สุด รายละเอียดของการคำนวณและผลการทดลองจะได้กล่าวในหัวข้อถัดไป

นอกจากค่าลักษณะสำคัญที่กล่าวมาแล้ว ยังได้มีการวิจัยที่ทดลองใช้ค่าลักษณะสำคัญแบบอื่นๆ อีก อาทิเช่น การประมาณพันธะแบบอิงการรับฟังของมนุษย์ (Perceptual linear predictive: PLP) ซึ่งยังไม่ได้ผลการระบุผู้พูดสูงนัก และการผสมค่าลักษณะสำคัญปกติกับค่าการเปลี่ยนแปลง (Derivative) ซึ่งแม้จะให้ผลดีกว่าแบบปกติก็จริง แต่ต้องใช้ลำดับของค่าลักษณะสำคัญจำนวนมาก เป็นการลดความเร็วของการประมวลผล

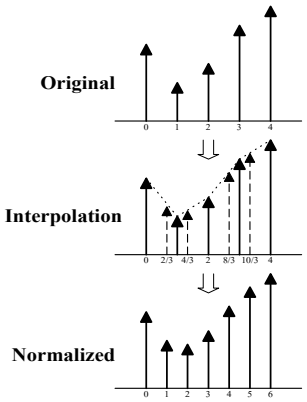
3.4 การทดลองอื่นๆ

นอกจากงานวิจัยในส่วนหลักที่กล่าวมาแล้ว ยังมีการวิจัยในรายละเอียดส่วนอื่นๆ ที่มีความสำคัญควรค่าแก่การพิจารณา เพื่อเพิ่มโอกาสในการพัฒนาผลการระบุผู้พูดให้ดีขึ้น ในที่นี้มีการวิจัยเพิ่มเติม 2 ส่วนดังนี้

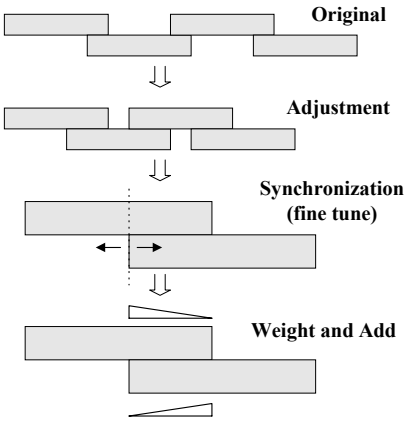
<u>เทคนิคการนอร์มอลไลซ์ทางเวลา</u>

ในขั้นตอนหนึ่งของการวิจัย ได้มีความพยายามปรับปรุงระบบรู้จำที่ใช้ ANN แบบปกติ เนื่องจากสามารถรู้จำได้โดยใช้เวลาประมวลผลไม่มากเมื่อเทียบกับ DTW สำหรับการพัฒนาระบบ ANN ดังที่ได้กล่าวมาแล้ว ส่วนสำคัญที่น่าจะเป็นตัวฉุดความสามารถของการรู้จำลง คือการนอร์มอลไลซ์ทางเวลา ซึ่งเป็นการปรับให้เสียงมีความยาวเท่ากันก่อนผ่านเข้าระบบ เพื่อให้ได้จำนวนข้อมูลที่จะส่งเข้า ANN เท่ากันทั้งหมด

วิธีการนอร์มอลไลซ์ทางเวลามีหลายวิธี ได้แก่ การเปลี่ยนอัตราการชักตัวอย่าง (Sampling rate changing) [4] การประมาณค่าในช่วงเชิงเส้น (Linear interpolation) [4] การเหลื่อมและรวมส่วนย่อยแบบซิงโครไนซ์ (Synchronized overlap-and-add: SOLA) [5] วิธีที่ดีจะต้องทำให้สัญญาณเสียงเพี้ยนไปจากเดิมน้อยที่สุดเท่าที่จะเป็นไปได้ จากการวิจัยที่ผ่านมาได้เปรียบเทียบ 2 วิธีคือ การประมาณค่าในช่วงเชิงเส้นและ SOLA หลักการของการนอร์มอลไลซ์ทางเวลาทั้ง 2 แบบแสดงในรูปที่ 5



ก. การประมาณค่าในช่วงเชิงเส้น



ข. SOLA

รูปที่ 5. การนอร์มอลไลซ์ทางเวลา 2 วิธี.

การประมาณค่าในช่วงเชิงเส้นเป็นวิธีที่ง่าย ทำโดยการเพิ่มหรือลดจำนวนเสียงของสัญญาณตัวอย่างให้มีขนาดตามต้องการ โดยเสียงสัญญาณใหม่จะถูกสร้างขึ้นจากสัญญาณเดิมสองข้างที่อยู่ติดกัน วิธีนี้จะทำให้สัญญาณเสียงเพี้ยนไป (Aliasing) แต่อย่างไรก็ตามระบบการเทียบเคียงสัญญาณเสียงก็ยังสามารถทำการรู้จำสัญญาณเสียงได้ ส่วนวิธี SOLA จะให้ความสำคัญในการปรับสัดส่วนลักษณะสำคัญของเสียงและคุณสมบัติทางเวลาให้มีความคล้ายคลึงกับสัญญาณเสียงต้นแบบมากที่สุด โดยการตัดสัญญาณเสียงเป็นช่วง ๆ แล้วนำมาซ้อนทับกัน ปรับเปลี่ยนระยะทางของสัญญาณที่นำมาซ้อนทับกัน โดยขึ้นอยู่กับสัดส่วนของเวลาที่ต้องการ ให้น้ำหนักความสำคัญของสัญญาณในแต่ละช่วงก่อนที่จะไปรวมกัน

การทดลองกับผู้พูดจำนวน 20 คนโดยใช้ค่าลักษณะสำคัญแบบ LPCC ขนาด 15 อันดับกับเสียงตัวเลขโดด 0-9 ผลการทดลองปรากฏว่าวิธี SOLA ให้อัตราการระบุผู้พูดเฉลี่ยถึง 75.33% [15] ในขณะที่วิธีที่ใช้การประมาณค่าในช่วงเชิงเส้นให้ผลเพียง 67.28 แสดงให้เห็นถึง ความสำคัญของการคงลักษณะดั้งเดิมของเสียงไว้ ดังนั้นถ้าเป็นไปได้ควรหลีกเลี่ยงการนอร์มอลไลซ์ทางเวลา

ผลกระทบของระดับเสียงของคำพูดที่ใช้ระบุผู้พูด

เสียงของภาษาไทยมีความแตกต่างจากเสียงภาษาต่างประเทศอยู่หลายประการ จุดสำคัญจุดหนึ่งคือ ภาษาไทยมีระดับเสียง (Tone) โดยมีห้าระดับคือ สามัญ (Middle) เอก (Low) โท (Falling) ตรี (High) จัตวา (Rising) งานวิจัยอีกส่วนหนึ่งคือการเปรียบเทียบผลของการใช้เสียงพูดในระดับเสียงต่างๆ ในการระบุผู้พูด โดยมีเป้าหมายเพื่อคัดเลือกคำที่เหมาะสมมาใช้ในการระบุผู้พูด

การทดลองระบุผู้พูดจำนวน 9 คน โดยใช้ค่าลักษณะสำคัญแบบ LPC ขนาด 10 อันดับและระบบรู้จำแบบ ANN คำพูดที่ใช้ในการระบุผู้พูดประกอบด้วย 6 ประโยคคือ “เออเออเออ”, “เอ่เอ่เอ่เอ่”, “เอ้เอ้เอ้เอ้”, “เอ๊เอ๊เอ๊เอ๊”, “เอ๋เอ๋เอ๋เอ๋เอ๋” และชุดสุดท้ายเป็นวรรณยุกต์ผสมคือ “เออเอ่เอ้เอ๊เอ๋” ผลการทดลอง [16] พบว่าชุดที่เป็นเสียงวรรณยุกต์ผสมให้ผลการระบุผู้พูดสูงที่สุดคือ 95.56% ส่วนวรรณยุกต์เดี่ยวๆ พบว่าเสียงวรรณยุกต์ในกลุ่มที่มีการเปลี่ยนแปลงในพยางค์คือวรรณยุกต์โทและจัตวาจะให้ผลได้ดีกว่าชุดอื่นๆ

4. ระบบระบุผู้พูดในปัจจุบัน

ณ ปัจจุบันนี้ ผลการทดลองได้ผลดีที่สุดถึง 92.30% สำหรับคำพูดตัวเลขโดด และเพิ่มขึ้นสูงกว่า 98% เมื่อใช้เสียงของตัวเลขต่อกัน ระบบดังกล่าวมีรายละเอียดดังต่อไปนี้

4.1 สัญญาณเสียง

ในงานวิจัย ได้ทำการอัดเก็บเสียงในรูปแบบของสัญญาณดิจิตอล โดยผ่านไมโครโฟนที่ต่อกับคอมพิวเตอร์ผ่านทางการ์ดเสียงปกติ กำหนดให้เก็บเสียงในรูปแบบของไฟล์ WAV อัตราการชักตัวอย่าง (Sampling rate) ที่ 11.025 กิโลเฮิร์ต ตัวอย่างละ 16 บิต และแบบช่องสัญญาณเดียว (Mono)

ทำการอัดเสียงจากผู้พูดจำนวน 50 คน (ชาย 30 คนและหญิง 20 คน) เป็นเวลา 5 สัปดาห์ ในแต่ละสัปดาห์ ผู้พูดแต่ละคนจะต้องพูดเสียงตัวเลขโดด 0-9 เป็นภาษาไทย ตัวเลขละ 10 ครั้ง และเพื่อป้องกันการที่ผู้พูดชินกับการพูดตัวเลขต่อๆ กัน จึงมีการพัฒนาโปรแกรมเฉพาะสำหรับการอัด โปรแกรมจะทำการสุ่มตัวเลขโดด 0-9 ขึ้นแสดงบนจอคอมพิวเตอร์ทีละตัว ผู้พูดจะต้องพูดเสียงตัวเลขที่แสดงเท่านั้น โปรแกรมจะกำหนดช่วงเวลาของการพูดไว้ไม่เกิน 1 วินาทีต่อหนึ่งตัวเลข หากผู้พูดคนใดพูดก่อนช่วงเวลาที่กำหนดหรือพูดช้ากว่าเวลาที่กำหนด โปรแกรมจะสั่งให้พูดใหม่อีกครั้งโดยอัตโนมัติ โดยพิจารณาจากค่าพลังงานของสัญญาณในช่วงเวลา 1 วินาทีดังกล่าว หลังจากนั้นจะแบ่งสัญญาณเสียงออกเป็น 2

กลุ่ม สัญญาณเสียงในสัปดาห์ที่ 1-3 ใช้สำหรับเป็นชุดฝึกฝนหรือชุดอ้างอิง ส่วนสัปดาห์ที่ 4-5 ใช้เป็นชุดทดสอบ

4.2 การประมวลผลขั้นต้น

กระบวนการประมวลผลสัญญาณขั้นต้นประกอบด้วย การกรองสัญญาณ (Filtering) โดยใช้ตัวกรองแบบดิจิตอลชนิดผ่านความถี่สูง กำหนดจุดผ่านความถี่ (Cutoff frequency) ที่ 200 เฮิร์ต เพื่อป้องกันสัญญาณรบกวนความถี่ต่ำที่เกิดจากแหล่งกำเนิดไฟฟ้า ต่อจากนั้นจะผ่านการตัดหัว-ท้ายเสียง (Endpoint detection) ในที่นี้อาศัยวิธีการตัดโดยพิจารณาจากค่าพลังงานของเสียง [4] โดยไม่มีการนอร์มอลไลซ์ทางเวลา

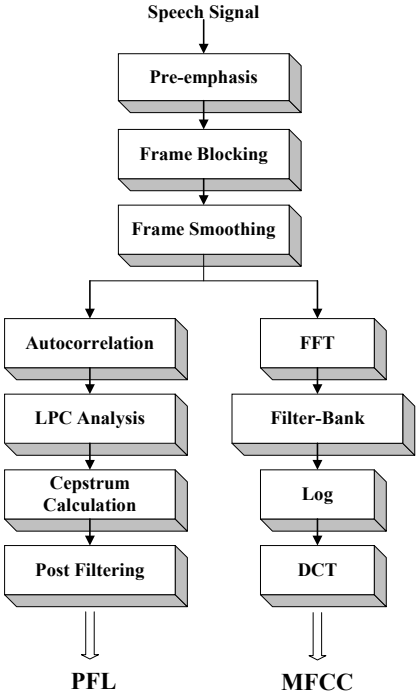
4.3 การสกัดค่าลักษณะสำคัญ

สัญญาณเสียงที่ผ่านการประมวลผลขั้นต้นแล้วจะนำมาผ่านการเน้นสัญญาณเบื้องต้น (Preemphasis) ด้วยตัวกรองอันดับที่ 1 (First order filter) เพื่อเพิ่มค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวน (Signal-to-noise ratio) หลังจากนั้นจะตัดแบ่งสัญญาณเสียงออกเป็นส่วนย่อย (Frame) ขนาดส่วนย่อยละ 20 มิลลิวินาที โดยเหลื่อมส่วนย่อยละ 5 มิลลิวินาที แต่ละส่วนย่อยจะผ่านการปรับให้ราบเรียบ (Smoothing) ด้วยแฮมมิงวินโดว์ (Hamming window) หลังจากนั้นจึงทำการสกัดค่าลักษณะสำคัญ ค่าลักษณะสำคัญที่ให้ผลการระบุผู้พูดสูงที่สุด ณ ปัจจุบันคือ MFCC และ PFL ขนาด 15 อันดับ ซึ่งมีรายละเอียดดังต่อไปนี้

MFCC [6] – ค่าสัมประสิทธิ์เซปสตรัมเป็นค่าลักษณะสำคัญที่นิยมมากทั้งในระบบรู้จำผู้พูดและเสียงพูด โดยพื้นฐานแล้วเซปสตรัมสามารถคำนวณได้จาก การแปลงโคซายน์แบบไม่ต่อเนื่อง (Discrete cosine transformation) ของค่าลอการิธึม (Logarithm) ของสเปกตรัม (Spectrum) ของสัญญาณเสียงแต่ละส่วนย่อย สเปกตรัมของสัญญาณเสียงสามารถหาได้โดยการแปลงฟูริเยร์แบบไม่ต่อเนื่อง (Discrete Fourior transformation) หรือการแปลงฟูริเยร์แบบเร็ว (Fast Fourior transformation) ขั้นตอนดังกล่าวตั้งอยู่บนพื้นฐานแนวคิดที่ว่า สเปกตรัมของสัญญาณเสียงกำเนิดจากส่วนประกอบ 2 ส่วนคือ เอนเวโลปของสเปกตรัม (Spectral envelop) และโครงสร้างรายละเอียดของสเปกตรัม (Spectral fine structure) ทั้ง 2 ส่วนสามารถแยกกันได้ด้วยการใส่ลอการิธึม สัมประสิทธิ์เซปสตรัมเป็นการแทนสัญญาณในส่วนเอนเวโลปของสเปกตรัมเท่านั้น

พัฒนาการหนึ่งของเซปสตรัม คือการผ่านสเปกตรัมของสัญญาณเสียงเข้าไปในกลุ่มของตัวกรอง (Filter bank) ซึ่งกระจายอยู่บนสเกลความถี่แบบไม่สม่ำเสมอ เช่น การกระจายตามสเกลเมล (Mel scale) [6] ซึ่งออกแบบมาให้เหมาะสมกับการรับฟังของหู เป็นต้น ค่าพลังงานของสเปกตรัมของเสียงที่ได้จากตัวกรองแต่ละตัวจะถูกนำมาใช้คำนวณค่าสัมประสิทธิ์เซปสตรัมแทนค่าสเปกตรัมปกติ ค่าสัมประสิทธิ์เซปสตัมที่ได้จากการกระทำเช่นนี้จึงได้ชื่อว่า MFCC

PFL [7,17] – อีกวิธีการหนึ่งของการคำนวณค่าสัมประสิทธิ์เซปสตรัมคือการคำนวณจากค่าสัมประสิทธิ์ LPC วิธีการคำนวณรวมทั้งเหตุผลของการคำนวณแสดงไว้ใน [3,18] หลังจากนั้นมีการเสนอค่าลักษณะสำคัญแบบใหม่ โดยผ่านค่าเซปสตรัมที่ได้เข้าไปยังตัวกรองซึ่งเรียกว่า ตัวกรองภายหลัง (Post filter) ตัวกรองดังกล่าวจะทำการเน้นค่าสเปกตรัมของเสียง ณ บริเวณความถี่ฟอร์แมนท์ (Formant frequency) ซึ่งเป็นการเพิ่มความโดดเด่นของสัญญาณเสียงทั้งในแง่การรู้จำเสียงพูดและรู้จำผู้พูด ภาพรวมของการคำนวณค่าลักษณะสำคัญทั้ง 2 วิธีที่กล่าวมาแสดงไว้ในรูปที่ 6

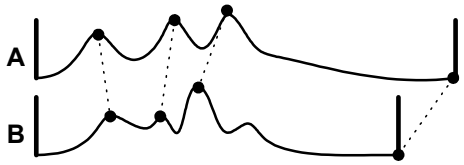


รูปที่ 6. ขั้นตอนการคำนวณค่าลักษณะสำคัญ

4.4 ระบบรู้จำ

ในปัจจุบันระบบรู้จำที่ให้ผลการระบุผู้พูดสูงสุดคือ DTW โดยใช้กฎการตัดสินใจแบบ K-NN รองลงมาคือ ANN โดยใช้วิธีป้อนข้อมูลเป็นช่วงของส่วนย่อย ส่วนต่อไปจะอธิบายหลักการคร่าวๆ ของแต่ละวิธี

DTW และ K-NN [6] – DTW เป็นวิธีการหนึ่งของการโปรแกรมพลวัต (Dynamic programming) ใช้ในการเทียบเคียงเพื่อหาระยะห่างระหว่างลำดับ 2 ชุดซึ่งยาวไม่เท่ากัน กำหนดให้ลำดับ A = $\{a_1, a_2, \ldots, a_I\}$ และ B = $\{b_1, b_2, \ldots, b_J\}$ เป็น 2 ลำดับที่จะเทียบเคียงกัน DTW จะทำการหาจุดเทียบเคียงที่ให้ค่าระยะห่างรวมต่ำที่สุด ดังแสดงในรูปที่ 7



รูปที่ 7. ภาพแสดงวิธีการเทียบเคียงแบบ DTW

โดยอาศัยขั้นตอนกระบวนการดังต่อไปนี้

ขั้นที่ 1: กำหนดค่าเริ่มต้น $D(a_1,b_1) = 2d(a_1,b_1)$ โดยที่ $d(a_i, b_j)$ เป็นค่าระยะห่างระหว่างจุด a_i และ b_j อาจใช้ระยะห่างแบบยูคลิเดียนก็ได้

ขั้นที่ 2: คำนวณแบบวนซ้ำหาจุดเทียบเคียง a_i และ b_j ที่เหมาะสมโดยตั้งอยู่บนพื้นฐานที่ว่า $D(a_i,b_j)$ จะต้องให้ค่าต่ำที่สุด ดังสมการ

$$D(a_i,b_j) = \min\begin{cases} D(a_{i-1},b_j) + d(a_i,b_j) \\ D(a_{i-1},b_{j-1}) + 2d(a_i,b_j) \\ D(a_i,b_{j-1}) + d(a_i,b_j) \end{cases} \qquad (1)$$

ทั้งนี้จะมีข้อกำหนดดังต่อไปนี้

1) $1 \le i \le I, 1 \le j \le J$
2) จุดเทียบเคียงจุดแรกคือ (a_1,b_1) และ (a_I, b_J) เป็นจุดเทียบเคียงจุดสุดท้าย
3) สำหรับแต่ละจุด (a_i, b_j) ที่เทียบเคียงกัน $|i - j| \le r$
4) $0 \le i^{k+1} - i^k \le 1, 0 \le j^{k+1} - j^k \le 1$ โดย k เป็นดัชนีรอบของการเทียบจุด

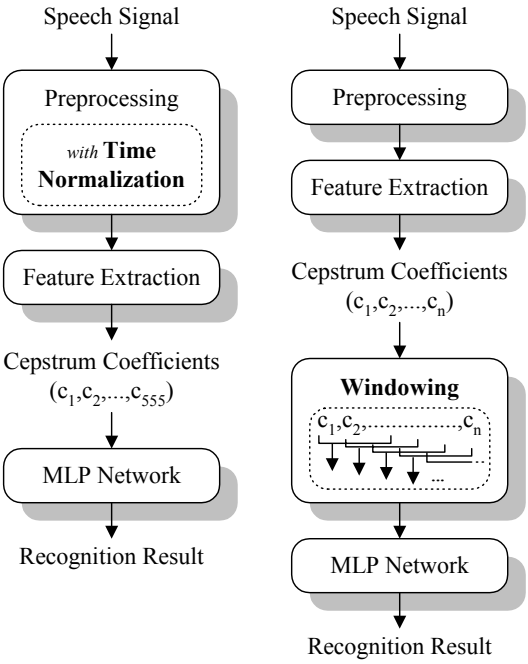
ขั้นที่ 3: ค่าระยะห่างรวมของ 2 ลำดับคือ $\frac{D(a_I, b_J)}{I + J}$

สมการที่ (1) จะตรงตามข้อกำหนดข้อที่ 4 ในตัวเอง กล่าวคืออนุญาตให้มีการขยับจุดที่จะเทียบเคียงได้ทีละ 1 จุดเท่านั้น ค่า r ในข้อกำหนดข้อที่ 3 เป็นค่าที่สำคัญที่ใช้ในการกำหนดความห่างของจุดที่เทียบเคียงกัน เพื่อให้บรรลุตามข้อกำหนดข้อที่ 2 จะมีการเติมเวกเตอร์ศูนย์ต่อท้ายลำดับที่สั้นกว่าดังที่ได้กล่าวมาแล้วในหัวข้อ 3.2 เพื่อให้สามารถจับคู่จุดสุดท้ายได้พอดี

เมื่อได้ค่าระยะห่างระหว่างสัญญาณเสียงที่เข้ามาทดสอบกับสัญญาณเสียงในชุดอ้างอิงแล้ว จะใช้วิธี K-NN ในการตัดสินใจ คือการพิจารณาสัญญาณเสียงอ้างอิง K ตัวที่ให้ค่าระยะห่างต่ำที่สุด ว่าไปตกลงที่สัญญาณเสียงของผู้พูดคนใดมากกว่ากัน ก็จะตอบเป็นผู้พูดคนนั้น ในการทดลองนี้ใช้ 5-NN ในการตัดสินใจ และจะเปลี่ยนเป็น 1-NN เมื่อ 5-NN ไม่สามารถตัดสินใจได้

ANN [8,19] – ANN ที่ใช้เป็นแบบ MLP และวิธีการเรียนรู้แบบแพร่กระจายย้อนกลับ โดยพัฒนาวิธีการป้อนข้อมูลเข้าแบบใหม่คือแบบหน้าต่าง เพื่อหลีกเลี่ยงการนอร์มอลไลซ์ทางเวลา หลักการที่พัฒนาขึ้นนี้เทียบกับวิธีการป้อนข้อมูลแบบเก่าได้แสดงไว้ในรูปที่ 8

ANN แบบเก่าจะมีจำนวนโหนดในชั้นข้อมูลเข้าเท่ากับ 555 โหนดคงที่ (15 ลำดับ * 37 ส่วนย่อยของเสียงที่ผ่านการนอร์มอลไลซ์ทางเวลามา) ส่วนจำนวนโหนดที่ชั้นข้อมูลออกจะเท่ากับจำนวนผู้พูดที่จะระบุและมีเพียง 1 โครงข่ายเท่านั้น แต่โครงสร้างแบบใหม่จะมี 1 โครงข่ายต่อผู้พูด 1 คน มีจำนวนโหนดในชั้นข้อมูลเข้าเพียง 60 โหนด (15 ลำดับ * 4 ส่วนย่อย) คือเลื่อนข้อมูลเข้าทีละ 4 ส่วนย่อยและเหลื่อมครั้งละ 3 ส่วนย่อย มีโหนดในชั้นข้อมูลออกเพียง 2 โหนด ซึ่งทำการรู้จำว่าใช่หรือไม่ใช่ผู้พูดคนนั้น



รูปที่ 8. วิธีการป้อนข้อมูลเข้า ANN

4.5 ผลการระบุผู้พูด

การทดลองระบุผู้พูดจำนวน 50 คนโดยอาศัยค่าลักษณะสำคัญและระบบรู้จำที่ได้กล่าวมาแล้ว แบ่งเป็น 2 ส่วน ส่วนแรกเป็นการระบุผู้พูดโดยใช้เสียงพูดตัวเลขโดด 0-9 ผลการทดลองแสดงไว้ในตารางที่ 2

ตารางที่ 2. ผลการระบุผู้พูดเมื่อใช้เสียงตัวเลขโดด

ตัวเลข	อัตราการระบุผู้พูด (%)			
	DTW		ANN	
	MFCC	PFL	MFCC	PFL
0	89.1	90.7	86.3	87.4
1	89.4	89.4	86.3	86.4
2	87.8	86.7	80.9	82.1
3	87.6	87.9	86.5	81.1

4	85.7	86.9	79.4	78.6
5	92.3	89.1	90.1	82.1
6	85.3	79.7	74.2	72.8
7	84.2	83.6	78.2	77.0
8	79.9	75.5	77.0	72.6
9	86.1	85.5	84.5	82.6
เฉลี่ย	86.74	85.50	82.34	80.27

ตารางที่ 3. ผลการระบุผู้พูดเมื่อใช้เสียง 3 ตัวเลขต่อกัน

ระบบ		ตัวเลข	ผลการระบุผู้พูด (%)
DTW	MFCC	“510”	98.70
	PFL	“015”	98.80
ANN	MFCC	“530”	97.30
	PFL	“019”	96.40

เพื่อเพิ่มประสิทธิภาพของการระบุผู้พูด จึงได้ใช้เสียงพูดที่ยาวขึ้นโดยการต่อเสียงตัวเลขโดดเป็น 3 ตัว ตัวเลขโดดที่นำมาต่อกันนั้นจะเลือกมาจากตัวเลขที่ให้ผลการระบุผู้พูดสูงที่สุด 3 อันดับแรกจากการทดลองกับเสียงตัวเลขโดด ผลการระบุผู้พูดรวมทั้งตัวเลขต่อกันที่ใช้ในการทดลองแสดงได้ดังตารางที่ 3

4.6 ซอฟต์แวร์ต้นแบบ

เมื่องานวิจัยมาถึงจุดที่ให้ผลการระบุผู้พูดที่สูง โดยมีอัตราการระบุผู้พูดสูงเกินกว่า 90% คณะนักวิจัยจึงได้พัฒนาซอฟต์แวร์ต้นแบบสำหรับระบุผู้พูด ใช้กับเสียงตัวเลขโดด โดยผู้ใช้สามารถเลือกได้ว่าจะใช้ DTW และ KNN หรือใช้ ANN ในการรู้จำ ทั้งนี้ซอฟต์แวร์จะกำหนดให้ผู้พูดแต่ละคนอัดเสียงตัวเลขที่กำหนดจำนวน 3 ครั้ง ระบบจะนำไปใช้เป็นชุดอ้างอิงสำหรับ DTW หรือนำไปฝึกฝนสำหรับ ANN ก่อนขั้นตอนการทดสอบระบุผู้พูดจริง ผู้สนใจสามารถติดต่อขอชมการทำงานของซอฟต์แวร์ต้นแบบได้ที่ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ

5. อุปสรรคและงานในอนาคต

งานวิจัยและพัฒนาระบบระบุผู้พูดสำหรับภาษาไทยได้ดำเนินการผ่านมา 1 ปีเต็มแล้ว พบอุปสรรคในการดำเนินงานอยู่หลายประการ ได้แก่

1. การรวบรวมข้อมูลจากการอัดเสียง เนื่องจากงานวิจัยมีวัตถุประสงค์ที่ต้องการสร้างระบบที่สามารถรู้จำผู้พูดที่มีประสิทธิภาพสูง ไม่ว่าเวลาจะผ่านไปนานเท่าไร ระบบควรจะสามารถระบุผู้พูดได้ในอัตราความถูกต้องใกล้เคียงเดิม ดังนั้นขบวนการจัดเก็บเสียงจึงจัดเก็บในหลายสัปดาห์ต่อเนื่องกัน ซึ่งทางคณะนักวิจัยได้ขอความร่วมมือจากพนังงานของศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติที่ประจำอยู่ ณ อาคารมหานครยิบซั่มช่วยสละเวลามาอัดเสียงให้ ปัจจุบันรวบรวมได้ 50 คน คณะนักวิจัยคาดหวังว่าจะเก็บเสียงได้อย่างน้อย 100 คนในอนาคต โดยอัดเสียงนักศึกษาฝึกงานและอาจจัดตั้งโครงการความร่วมมือกับองค์กรต่างๆ เพื่อพัฒนาฐานข้อมูลเสียงภาษาไทย
2. เวลาในการระบุผู้พูดของ DTW ค่อนข้างนานมากเมื่อเทียบกับ ANN แต่เนื่องจากอัตราการรู้จำผู้พูดของ DTW ดีกว่าผลจาก ANN คณะนักวิจัยจึงต้องคิดค้นหาวิธีปรับปรุงเทคนิค DTW ให้สามารถทำงานได้เร็วยิ่งขึ้นหรือพัฒนา ANN ให้สามารถรู้จำได้ถูกต้องมากขึ้น นอกจากนี้คณะนักวิจัยกำลังทดลองเทคนิคการรู้จำผู้พูดวิธีอื่นๆ ด้วย เพราะคาดว่าจะได้ระบบระบุผู้พูดที่ดีมีประสิทธิภาพสูงขึ้นกว่าระบบในปัจจุบัน
3. ฐานความรู้เกี่ยวกับเสียงภาษาไทยค่อนข้างมีจำกัด ส่งผลให้คณะนักวิจัยมีความจำเป็นต้องทดลองในทุกๆ สมมติฐานที่ตั้งขึ้นเอง ซึ่งทำให้ต้องใช้เวลาในการวิจัยมากขึ้น ดังนั้นเมื่อคณะนักวิจัยได้ผลการทดลองจึงได้พยายามจัดทำบทความวิชาการเพื่อเผยแพร่ความรู้อยู่ตลอดเวลา ด้วยความหวังที่ว่าความรู้เหล่านั้นจะได้ช่วยเสริมให้เกิดงานวิจัยทางด้านการระบุผู้พูดในประเทศไทยมากขึ้น และลดงานที่ซ้ำซ้อนลงไป

6. บทสรุป

ระบบระบุผู้พูดมีความสำคัญมากกับการเพิ่มประสิทธิภาพของระบบรักษาความปลอดภัย และยังเป็นพื้นฐานที่สำคัญในการพัฒนาระบบรู้จำเสียงพูด (Speech recognition system) สำหรับภาษาไทยอีกด้วย ในปัจจุบันคณะนักวิจัยได้พัฒนาซอฟต์แวร์ต้นแบบระบุผู้พูดสำหรับภาษาไทยที่ให้อัตราความถูกต้องเฉลี่ย 98% กับผู้พูดจำนวน 50 คน คณะนักวิจัยจะพัฒนาระบบนี้ให้สามารถทำงานได้ดีกับผู้พูดจำนวนมากยิ่งขึ้น และจะพัฒนาระบบระบุผู้พูดที่ใช้งานได้กับเสียงพูดผ่านทางสายโทรศัพท์ด้วย นอกจากนี้ทางคณะนักวิจัยกำลังพัฒนาและรวบรวมฐานข้อมูลเสียงพูดตัวเลข 0-9 เพื่อให้นักวิจัยจากที่อื่นๆ สามารถเข้ามาในเว็บไซต์และดึงข้อมูลเพื่อนำไปใช้ทดลองและคิดค้นเทคนิคใหม่ๆ เพื่องานระบุผู้พูดสำหรับภาษาไทยที่ดีขึ้นอีกด้วย

เอกสารอ้างอิง

[1] J. P. Campbell, Jr., "Prolog to Speaker Recognition: A Tutorial", Proceedings of IEEE, Vol. 85, No. 9, pp. 1437-1462, September 1997.

[2] G. R. Doddington, "Speaker Recognition-Identifying People by their Voices", *Proceedings of IEEE*, Vol. 73, No. 11, pp.1651-1664, November 1985.

[3] S. Furui, "Digital Speech Processing, Synthesis, and Recognition", New York and Basel: Marcel Dekker, Inc, 1989.

[4] ชัย วุฒิวิวัฒน์ชัย, "การรู้จำเสียงคำหลายพยางค์แบบไม่ขึ้นกับผู้พูด โดยใช้เทคนิคแบบฟัซซีและนิวรอลเน็ตเวอร์ก", วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย, 2540

[5] S. Roucos and A.M. Wilgus, "High Quality Time Scale Modification for Speech," *IEEE International Conference ASSP*, pp. 493-496, 1985.

[6] L. R. Rabiner and B. -H. Juang, "Fundamentals of Speech Recognition", A. Oppenheim, Series Editor, Englewood Cliffs, NJ: Prentice-Hall, 1993.

[7] R. J. Mammone, X. Zhang, and R. P. Ramachandran, "Robust Speaker Recognition, A Feature-based Approach", *IEEE Signal Processing Magazine*, p. 58-71, September 1996.

[8] L Fausette, "Fundamentals of Neural Networks–Architecture, Algorithm, and Applications", Prentice-Hall, 1994.

[9] K. Yu, J. Mason, and J. Oglesby, "Speaker Recognition using Hidden Markov Models, Dynamic Time Warping and Vector Quantisation*", IEE Proc.-Vis. Image Signal Process*, Vol. 142, No. 5, October 1995.

[10] วศิน สินธุภิญโญ, เปรมนาถ ดูเบ, สุทัศน์ แซ่ตั้ง, วารินทร์ อัจฉริยะกุลพร, ชัย วุฒิวิวัฒน์ชัย และจุฬารัตน์ ตันประเสริฐ, "การระบุผู้พูดด้วย LPC และ DTW สำหรับภาษาไทย", *เอกสารประกอบการประชุมวิชาการ ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีประจำปีงบประมาณ 2542 ณ ศูนย์ประชุมสหประชาชาติ 30 มีนาคม-เมษายน 2542*

[11] S. Sae-Tung and C. Tanprasert, "Feature Windowing based Thai Text-Dependent Speaker Identification using MLP and Backpropagation Algorithm*", Proceedings of the International Symposium on Circuits and Systems*, May 2000.

[12] C. Wutiwiwatchai, V. Achariyakulporn, and C. Tanprasert, "Text-dependent Speaker Identification using LPC and DTW for Thai Language", *1999 IEEE 10th Region Conference (TENCON'99)*, Vol. 1, September 1999.

[13] วารินทร์ อัจฉริยะกุลพร, ชัย วุฒิวิวัฒน์ชัย และ จุฬารัตน์ ตันประเสริฐ, "ระบบระบุผู้พูดภาษาไทยด้วยวิธีไดนามิกส์ไทม์วาร์ปปิง", *กำลังพิจารณาเพื่อตีพิมพ์ใน NECTEC Technical Journal*, Vol.2, No. 7, 2000

[14] C. Wutiwiwatchai and C. Tanprasert, "Thai Text-Dependent Speaker Identification: Features Comparison", *The 4th Symposium on National Language Processing*, May 2000.

[15] C. Wutiwiwatchai, S. Sae-Tung, and C. Tanprasert, "Thai Text-Dependent Speaker Identification by ANN with Two Time Normalization Techniques", *Proceedings of the 1st Workshop on Natural Language Processing and Neural Networks*, pp. 47-52, November 1999.

[16] C. Tanprasert, C. Wutiwiwatchai, and S. Sae-tang, "Text-dependent Speaker Identification Using Neural Network on Distinctive Thai Tone Marks", Proceedings of *International Joint Conference on Neural Networks*, July 1999.

[17] M. S. Zilovic, R. P. Ramachadran, and R. J. Mammone, "Speaker Identification Based on the Use of Robust Cepstral Features Obtained from Pole-Zero Transfer Functions", *IEEE Transactions on Speech and Audio Processing*, Vol.6, No.3, pp.260-267, May 1998.

[18] S. Furui, "Cepstral Analysis Technique for Automatic Speaker Verification", *IEEE Transaction on Acoustic, Speech Signal Processing*, Vol. ASSP-29, pp.254-272, April 1981.

[19] SNNS (Stuttgart Neural Network Simulator) User Manual, Version 4.1, University of Stuttgart, Institute for Parallel and Distributed High Performance Systems (IPVR), Report No. 6/95.

นายชัย วุฒิวิวัฒน์ชัย ตำแหน่งผู้ช่วยนักวิจัย ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ สำเร็จการศึกษา ปริญญาโท (วิศวกรรมศาสตร์มหาบัณฑิต) ปี 2540 จุฬาลงกรณ์มหาวิทยาลัย ประสบการณ์การทำงาน โครงการระบบรู้จำเสียงพูดภาษาไทยซึ่งอยู่ในช่วงเริ่มต้น ผลงานเด่น มีชื่อเป็นผู้แต่งบทความวิชาการที่ได้รับตีพิมพ์ภายในประเทศจำนวน 1 บทความ และระดับนานาชาติ จำนวน 7 บทความ

วารินทร์ อัจฉริยะกุลพร สำเร็จการศึกษาระดับปริญญาตรีสาขาวิทยาการคอมพิวเตอร์ เกียรตินิยมอันดับ 2 จากมหาวิทยาลัยเกษตรศาสตร์ ปี พ.ศ. 2538 และการศึกษาระดับปริญญาโทสาขาวิทยาการคอมพิวเตอร์ จากมหาวิทยาลัยมหิดล ปี พ.ศ. 2541 ปัจจุบันได้ร่วมงานในหน่วยปฏิบัติการวิจัยและพัฒนาวิศวกรรมภาษาและซอฟต์แวร์ โดยรับผิดชอบในหน้าที่การพัฒนาโปรแกรมพัฒนาเว็บเพจ และฐานข้อมูล นอกจากนี้ยังมีส่วนร่วมรับผิดชอบในโครงการแก้ปัญหาคอมพิวเตอร์ปี ค.ศ. 2000 โครงการ Asean-India Digital Archive (AIDA) และโครงการระบบระบุผู้พูดภายใน NECTEC มีความสนใจในด้านการออกแบบระบบงาน การออกแบบโปรแกรมเชิงวัตถุ และงานทางด้านปัญญาประดิษฐ์

สุทัศน์ แซ่ตั้ง สำเร็จการศึกษาระดับปริญญาตรีสาขาระบบสารสนเทศ เกียรตินิยมอันดับ 2 จากสถาบันเทคโนโลยีรัชมงคล ปี พ.ศ. 2536 และการศึกษาระดับปริญญาโทสาขาวิทยาการคอมพิวเตอร์ จากมหาวิทยาลัยมหิดล ปี พ.ศ. 2541 ปัจจุบันได้ร่วมงานในหน่วยปฏิบัติการวิจัยและพัฒนาวิศวกรรมภาษาและซอฟต์แวร์ โดยรับผิดชอบในงานวิจัย และพัฒนาโปรแกรมแปลงรูปภาพเอกสารเป็นข้อความ (Thai OCR) โครงการระบบระบุผู้พูดด้วยเสียง นอกจากนี้ยังมีส่วนร่วมรับผิดชอบในโครงการแก้ปัญหาคอมพิวเตอร์ปี ค.ศ. 2000 มีความสนใจในด้านการประมวลรูปภาพ (Image Procesisng) การรู้จำรูปแบบ (Pattern Recognition) และการออกแบบพัฒนาโปรแกรมเชิงวัตถุ (Object Oriented Systems)

Issues in Thai Text-to-Speech Synthesis: The NECTEC Approach[1]

Pradit Mittrapiyanuruk, Chatchawarn Hansakunbuntheung,
Virongrong Tesprasit and Virach Sornlertlamvanich
Information R&D Division,
National Electronics and Computer Technology Center (NECTEC)
Gypsum Metropolitan Building, 22nd Floor,
539/2 Sri Ayudhaya Road, Rajthevi, Bangkok 10400, Thailand
(pmittrap, chatchawarnh)@notes.nectec.or.th, (virong, virach)@nectec.or.th

ABSTRACT – This paper presents all the essential issues in developing the text-to-speech synthesis for Thai - text analysis, prosody generation and speech synthesis. In the text analysis, problems in Thai text processing can be decomposed into the models of sentence extraction, phrase boundary determination and grapheme-to-phoneme conversion. The syllable duration and F0 contour generation rules are included in the prosody generation. This is to realize the synthetic speech in the suprasegmental level. In the speech synthesis, the definition and the construction of acoustic inventory structure 'demisyllable' are presented. Furthermore, three signal-processing algorithms, amplitude normalization, the segment boundary smoothing and prosodic modification, are also presented in this topic.

KEY WORDS -- Thai text-to-speech synthesis, text analysis, prosody generation, speech synthesis, demisyllable

บทคัดย่อ -- บทความนี้นำเสนอหัวข้อสำคัญในการวิจัยและพัฒนาระบบสังเคราะห์เสียงพูดจากข้อความภาษาไทย ประกอบด้วยการวิเคราะห์ข้อความ, การสังเคราะห์สัทสัมพันธ์และการสังเคราะห์สัญญาณเสียงพูด ในหัวข้อการวิเคราะห์ข้อความจะกล่าวถึงปัญหาที่สำคัญในการประมวลผลข้อความภาษาไทยและรายละเอียดของส่วนประกอบภายในซึ่งประกอบด้วย 3 ส่วน ได้แก่ การตัดประโยค การหาขอบเขตวลีเพื่อหยุดเว้นวรรคการอ่าน และการแปลงรูปเขียนเป็นรูปเสียงอ่าน ในหัวข้อการสังเคราะห์สัทสัมพันธ์จะกล่าวถึงกฎในการกำหนดช่วงเวลาของพยางค์และ F0 contour ซึ่งจะทำให้สามารถสังเคราะห์เสียงที่มีความสัมพันธ์ในระดับเหนือหน่วยเสียงได้ ส่วนหัวข้อการสังเคราะห์สัญญาณเสียงพูดจะกล่าวถึงโครงสร้างหน่วยเสียงแบบครึ่งพยางค์และอัลกอริทึมทางการประมวลผลสัญญาณในการปรับสัญญาณที่รอยต่อให้ต่อเนื่องและปรับสัญญาณให้มีสัทสัมพันธ์ตามที่ได้กำหนดมา

คำสำคัญ -- การสังเคราะห์เสียงพูดจากข้อความภาษาไทย, การวิเคราะห์ข้อความ, การสังเคราะห์สัทสัมพันธ์, การสังเคราะห์สัญญาณเสียงพูด, ครึ่งพยางค์

1. Introduction

Text-to-speech synthesis is a module or system or machine that converts the input text into the acoustic speech signal that people can understand. Many kinds of applications utilized from this system are developed such as the applications for blind people e.g. screen reader, or the applications for normal people e.g. electronic mail reader using telephone interface, etc. Most of the text-to-speech synthesis systems are developed for converting the text for major languages such as English, Chinese, Japanese and the European languages. At the present, there are only few

[1] This article is a reprint of the article appeared in the Proceedings of NECTEC Annual Conference 2000 : ECTI Technologies for New Economies, June 2000, pp. 483-495. This paper wins a best paper award in category of "Best Presentation".

systems developed for the Thai language. Most of them lack for the continuity in their milestone and some focus on the specific point rather than the whole picture. As a result, there is no Thai text-to-speech synthesis system using in the real application. To overcome this obstacle, this work attempts to put together the jigsaw to form a complete picture. Our goal is to produce a text-to-speech synthesis that can synthesize a natural sound.

In retrospect, there are some research works related to Thai text-to-speech synthesis. The Luksaneeyanawin's system [1] consists of three main modules. First is the Thai text processing module, it converts a string of Thai text into a string of Thai phonological units using the syllable, word and phrase parsers. Second is the sound dictionary module. It looks up the synthesis unit for the corresponding phonological unit. Third is the synthesis by waveform concatenation module. It synthesizes the speech by using the waveform concatenation technique. Taisertavattanakul and Kanawaree [2] developed a simple but practical system. The system contains (1) the text-to-phoneme analyzer by using conversion rules and a small dictionary for exceptional words (2) the synthesizer concatenates the speech waveform from the demisyllable based acoustic inventory.

The other research works that focus on some specific points rather than the whole system are [3, 4, 5]. Kiat-arpakul, Fakcharoenphol and Keretho [3] proposes an acoustic inventory structure for Thai speech synthesis. In this work, a syllable waveform is created from the concatenation of the phoneme-based and the demisyllable-based units. Luksaneeyanawin [4] proposes a technique to transform the tonal patterns of any syllable speech units by PSOLA-based resynthesizing F0 contour. This technique takes the advantages in the reducing the number of synthesis units about 5 times. This technique stores only toneme syllabic units and synthesizes other toneme speech from these toneme units. Hansakunbuntheung [5] applies the line spectrum pair to the Thai syllabic speech synthesizer. The sound units are encoded in the form of the 20^{th} order LSP and its residues. The synthesizer can synthesize all Thai five tones and adjust the sound duration by using the pitch-synchronous overlap-add (PSOLA) technique. The details of the literature survey in the field of Thai text-to-speech synthesis can be found in Luksaneeyanawin's work [6].

In our work, the system is divided into 3 major parts: text analysis, prosody generation and speech synthesis. The main function of text analysis is to segment the input text into smaller units: sentences and phrases, and then transcribe into the phoneme description. The prosody generation then determines the prosody parameter from the information analyzed by the text analysis. The phoneme description with the prosody parameter of the text is synthesized to the speech waveform by the speech synthesis module. In this module, any synthetic speech is created by the concatenative technique based on the demisyllable units. The signal processing algorithms are involved to produce the natural synthetic speech.

In this paper, the detailed of NECTEC's Thai text-to-speech synthesis is discussed. Most parts of its are already implemented. However, they are being improved in the naturalness. Section 2 discusses the issues in text analysis. The prosody generation based on the rewriting rules is discussed in Section 3. The detail in the acoustic inventory structure and signal processing algorithm are discussed in the speech synthesis topic, Section 4.

2. Issues in Text Analysis

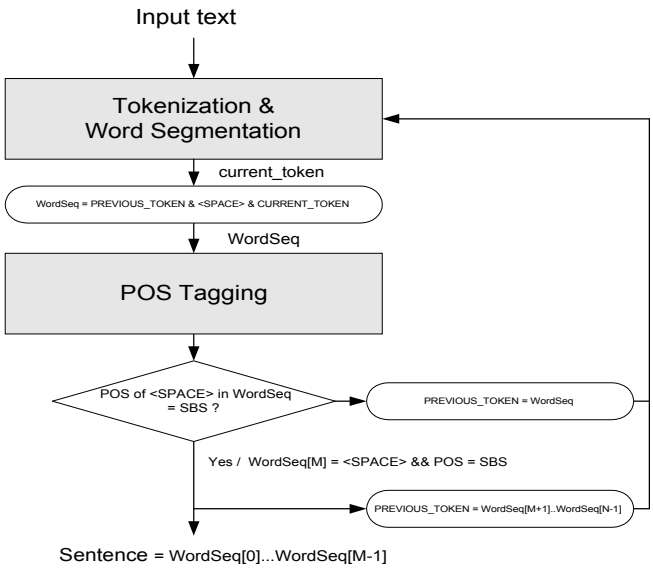
The text analysis is the first part to accept the input text into the system. In practice, the input text to the system may be one or more text paragraphs. Each paragraph consists of sentences. The text may include Thai words, foreign texts (e.g. English), and other special expressions such as numeral texts, abbreviations, punctuation marks, etc. Because the aim of this work is to synthesize the speech of Thai text, the foreign text that appears in the input text will be ignored. It is impractical to process the whole input text all at once due to the limitation in memory resource and processing time. Therefore the text analysis segment the input text into smaller units for processing in other modules. In this work, the text analysis will segment the input text into a sequence of sentences. In the same time, it also determines the phrase boundary, the acceptable position to pause when reading, for synthesizing a natural sound. In addition to determination the sentence and phrase boundary, a module called 'graphme-to-phoneme' in the text analysis is also included. It converts the text into the phonological representation.

2.1 Sentence Extraction

Unlike the English or other European languages, there is no explicit sentence marker in the Thai language. It is convention to insert the space at the end of a sentence in Thai writing. But not all spaces in a paragraph are the sentence marker. They also can be used as other purposes [7,8] such as, using between phrases or cluases within a sentence, between sentences in a cohesive group of sentences, before and after numerals, etc. Mittrapiyanuruk's and Sornlertlamvanich's work [9] extended the algorithm for POS tagging in probabilistic n-gram model to discriminate the sentence break spaces from other purpose spaces. The task can be view as the classification problem. We define the space by its function into 2 different types: sentence break and non-sentence-break space and apply the statistical part-of-speech (POS) tagging as the classifier.

The block diagram of sentence extraction algorithm is shown in Figure 1. The tokenization and word segmentation stage extract a set of tokens with at least one space in between. The spaces in the set of tokens are classified by POS tagging. A token is a sequence of consecutive characters enclosed by the spaces. The special expressions e.g. numerals, abbreviations, punctuation marks, etc. are specially considered. For example, '10600' can be pronounced either in the form of digit-by-digit reading as in the phrase of 'กทม. 10600', or in the form of quantity number as in the phrase of '10600 บาท'. The normalization is needed in this process.

Unlike the approaches used for English text, we do not expand this special expression into a normal text. It is embedded in the module of tokenization and grapheme-to-phoneme conversion. The tokenization module draws out this kind of text as a single token. Then the grapheme-to-phoneme conversion assigns a different phonological representation according to the text feature. Moreover, the



string of characters is segmented into words because there is no explicit use of word delimiter in general Thai text [10].

Figure 1. Block diagram of sentence extraction

The two adjacent tokens are reconstructed to the word sequence with a space in between. Any spaces in this word sequence are classified to be one of two possible classes, sentence break or non-sentence-break space. We define this classification problem in terms of statistical POS tagging. The most probable sequence of POSs and individual word-level POS assignments determines the most probable POS assignment of any word sequences. Therefore the classification task is to determine whether the POS of any spaces in the most probable sequence of POS is a sentence-break or not. We use the part-of-speech trigram model to compute the POS sequence probabilities and introduce the viterbi algorithm for computing the most probable sequence of POSs. Because it is possible that the space in the word sequence that used to be the non-sentence-break spaces are incorrectly assigned. Therefore we must scan the space between the current and previous token as well as the spaces within the previous token. If there is no sentence-break space then all of this word sequence will be used as the previous token in the next iteration. But if a sentence-break space is found then the output sentence is the whole sequence from the first word to the word just before the sentence-break space. The rest word sequence after the space is used as the first token in the next iteration. The algorithm will extract the sequence of tokens and detect the type of in-between space until the end of paragraph. It is obvious that this algorithm can solve the problem of memory resource and processing time limitations because it processes token by token rather the whole paragraph at once.

2.2 Phrase boundary determination

Normally when speakers utter a long sentence, they tend to break the sentence up into several phrases. Luksaneeyanawin [1] reports that from the study of charactersics and function of pause, the position of pause occurs averagely at each 8 syllables (S.D.=4). These phrase boundaries capture several prosodic characteristics, such as pausing at the end of phrase, lenthening in duration of phrase final syllable and the downtrend effect on F0 contour at the end of phrase, etc. If the phrase boundary is determined accurately then the prosody generation will undoubtedly play an important role in obtaining more natural speech. To accomplish the phrase boundary determination, the rule-based algorithm is developed using the output from the sentence extraction. The algorithm first determines the tentative phrase break positions in a sentence and then merges the positions using the syllable number constraint.

The algorithm places a tentative phrase break at the position of space and punctuation mark. The algorithm detects the phrase break spaces by using distinctive pattern rules derived from formal Thai writing pattern [11]. The other tentative phrase break positions are determined by scanning each pair of words from left to right in the sentence. The simple rule based on the content word/function word [12] is used to place the preliminary break position before every function word that follows a content word. In this work, the function word is a word of conjunction, preposition and relative pronoun. The content word is the rest that is not match the function word. Furthermore, the rule derived from Luksaneeyanawin's work [1] assigns the break position before/after some specific grammatical word.

At the merging step, the tentative phrase break positions in a sentence are combined to a single phrase if they do not end in the phrase-break space or punctuation and contain 10 or fewer syllables. They are combined to the following phrase until a punctuation mark or phrase-break space is found or until the number of syllables is greater than 10. This scheme of phrase combining is same as Karn's [13].

2.3 Grapheme-to-phoneme conversion

Outputs from the sentence extraction and phrase boundary determination are one sentence with the phrase break positions and the word boundaries. In this step, the phonological representation or phoneme of each word is assigned. First each word is looked up for the phoneme string from the pronunciation dictionary, about 25,000 entries. The letter-to-sound rule is developed to handle the unregistered words. The rule consists of 2 stages. First the grapheme of a word is divided into a syllable sequence and second the syllable sequence is converted to a phoneme string by simple orthographical-to-phonological mapping. The tone of each syllable is assigned by considering the phonological composition (initial consonant, vowel, final consonant) and its orthographical tone marker. Details in the

tone assigment rule can be founded in Thavaranon's work [8].

The first step typically called syllabification is developed using the regular expression. Rather than hard coding the rule for each syllable pattern, this work rewrites the rule in the regular expression format. All possible orthographical syllablic structure are listed in the regular expression format and compiled to be a deterministic finite state automata by the lexical analyzer or 'LEX'. When it matches a syllabic pattern then the orthographical syllablic composition: initial consonant, vowel final consonant and tone marker, is returned for assigning the phonological representation. The advantages of this scheme are the flexibility in rule modification and the speed of processing time.

3. Issues in Prosody Generation

The prosody means the properties of the acoustical speech such as pitch variation, loudness and syllable length. The effects of prosody are referred to as suprasegmental phenomena [14], since it occurs in higher level than segmental level such as syllable or phoneme. It is acknowledgable by most researchers in this field that the naturalness of synthetic speech is considerably affected by the prosody. Therefore this work essentially includes the prosody generation. Many prosody parameters are generated by determining the pause position and duration and the pitch movement of utterance which represented by F0 contour. The pausing is executed by the phrase boundary determination in text analysis part.

There are two major approaches in the research of prosody generation: the rule-based method and the corpus-based method. In the rule-based approach, linguistic experts derive the factors that affect the prosody event by observing various phenomena in the natural speech, then write the rules that interplay among these factors for synthesizing the more natural speech. On the contrary, the corpus-based approach derives the prosody model from the prosodic annotated speech corpus by using machine learning algorithms such as decision tree, artificial neural network, etc. The prosodic parameters of unseen text are determined by infering from the training corpus. Lacking of the prosodic-labeled speech corpus, our prosody generation is a rule-based approach.

3.1 Syllable duration assigment rule

The first consideration when devise the durational rule is the choice of speech unit that will be affected by the rule. The contexual influences that affected the duration of different speech units are varied. Campbell and Isard [15] argue that the syllable is a suitable unit that reflects the rhythm of any utterances. This approach first predicts the syllable duration and then the smaller segment duration e.g. phone is determined from its syllable duration. Because the speech synthesizer that we use in this work is the demisyllable-based concatenative system, the speech waveform is formed by the sequence of syllabic sounds. Each syllablic waveform is created by concatenating two demisyllable units: initial and final unit. Then the timing of synthetic speech appears in the syllabic time frame. There are many linguistic works [16] conclude that Thai is a syllable-time rhythm in which the syllable is an intuitively recognizable unit for primitive people. Therefore we select the syllable as the speech unit for modeling the duration.

In timing aspect, the naturalness of any utterance occurs when the duration of every syllable in the phrase is relatively suitable. In any slow and fast utterance, the duration of syllable differs only in the absolute value but the relative value is almost the same. To accomplish this task we tailor the most favorite scheme [17] to the prosodic generation module in the syllabic framework. This scheme first assigns the base syllable duration from its intrinsic property. Then the rules are used to multiply the base duration by a specific factor. These factors are devised by investigating the natural speech in word, phrase and sentence level. In this scheme we can adjust the speaking rate by multiplying the factor to the intrisic duration without changing the rule.

For finding the intrinsic duration of each syllable, it is laborious to acquire the intrinsic duration of all Thai syllablic sounds because of its plentitude of units, which is about 27,000 [6]. To realize this process, we classify the consonants by the manner of articulation into 8 types and the vowels by the tongue advancement/short/long attribute into 12 types and use mid tone (tone 0). By the assumption that the syllables in the same group have the same intrinsic duration, we use the duration of each group representative as the duration of every syllable in the group. This method reduces the number of syllable duration patterns to 384 patterns. The duration of each unit is taken from the carrier syllable in the medial position of pronunciation. The intrinsic interval of all syllables is extrapolated using the value of the representative that has the same kind of initial consonant, final consonant and vowel. The duration of falling (tone 2) and rising (tone 4) tonal syllable is scaled-up by factor 4/3 to compensate the tonal-durational interactive effect. Moreover we also measure intrinsic pause duration which is divided into 3 types: pre-plosive pause, glottal closure pause and end of phrase pause in the same way.

For the details of rule, they are derived from Klatt's work [17]. The rule in phrase and sentence level are the same which is lengthening the phrase-final syllable duration by the factor of 1.2 and inserting the pause at the end of phrase with the intrinsic phrase pause duration. At the end of sentence, the pause duration is longer than the intrinsic phrase pause duration by the factor of 1.2. For the rule in word level, the syllables in any words that are not the non word-final position are shortened by the factor of 0.9. Other syllable, any syllables in a polysyllabic word are shortened by the factor of 0.9. The last rule considers the effect of postvocalic consonant context. It shortens the duration of syllables followed by the voiceless consonant by the factor of 0.9. Noted that these rules apply sequentially by cumulatively multiplying the initial duration with each specfied rule's factor to obtain the final duration.

3.2 F0 contour Generation rule

In natural speech, the speech is continuously uttered as strings of breaths. Each string, called phrase, consists of

many sound units. The types of sound unit can be words, syllables or phonemes, etc. depending on the design purpose. Considering a particular type of the sound units, since the sound units in a phrase are produced in the same utterance, they must share some common characteristics. A characteristic, called intonation, is in the suprasegmental level of speech, relates to the tonal phenomena that affect on F0 contour of the continuous speech. In addition, there is another effect on syllabic level. There is a tone pattern when syllables are connected.

3.2.1 Intonation Rules

In the suprasegmental level, two groups of rules are defined. The first group is the downdrift phenomenon that defines how the F0 contour decreases relatively with the preceding time. Another group concerns the pitch range of F0 contour that limits the boundary of F0 contour.

3.2.1.1 Downdrift

A downdrift, can be observed in the F0 contour across a phrase [11,18-24]. [25] shows that this downdrift also happens in Thai speech as shown in Figure 2. This phenomenon can be observed by plotting the F0 contour of a phrase containing only the mid tone (tone 0). The plot shows the downdrift effect on the F0 contour which looks like the steps of similar patterns. The reference line connects all the beginning points of F0 contours of the syllables.

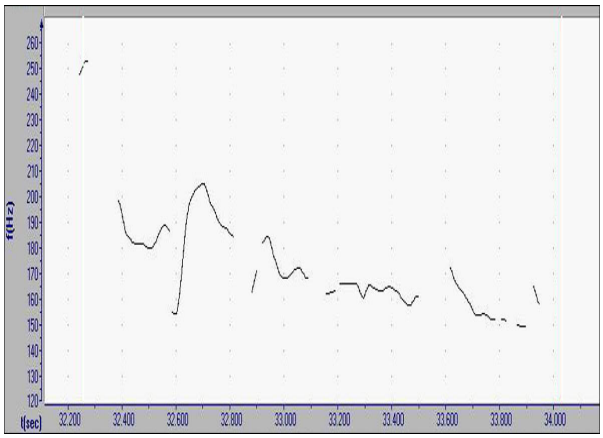


Figure 2. An example of downdrift on F0 contour of Thai speech

To simplify the effect, the downdrift is estimated by a linear declining slope of the F0 contour. Since one system is for a female speech, the slope was computed for a female speech prototype. As a result of the experiment, the declination is 30 hertz per second. This declination can be represented by a linear line, called a reference line, as shown in Figure 3. This line is used as the reference of the starting or ending point of F0 contour of each tone.

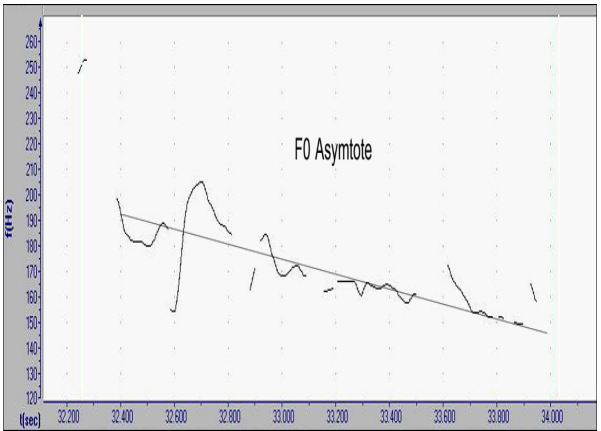


Figure 3 A reference line that expresses a simiplified version of downdrift

3.2.1.2 Pitch Range

An obvious difference between male and female speech is the pitch range. Generally, a female speech is more perceptible higher than a male speech. The pitch range specifies how high and how low the pitch level can reach.

To determine a pitch range, an observation on a female's speech prototype is done by measuring the maxima and minima of F0 levels. The boundaries of pitch range are applied to the system to limit the level of synthetic F0 contour with downdrift effect. If F0 value of syllabic F0 contour exceed these limits, the level of the syllabic F0 contour will be reset to the starting F0 of the same phrase.

3.2.2 Tone Rules

After processing on the suprasegmental level, here, a syllabic level will be discussed. In this level, there are two parts. The first part explains where the tone contours should be located and another part explaining the effect of adjacent syllables on a tone contour.

3.2.2.1 Tonal contour location

This part explains how to locate the tone contours. When tone contours are concatenated, the locations of tone contours are different depending on the situation. In a reading speech, the Mid, Low and Rise tone start at the reference line while the others end at the reference line. However, there are some special cases that the tone contours do not conform to this rule. If there is a stress syllable in a phrase, its tone contour level will be shifted up. Since only the reading speech synthesizing is the goal of this system, all tone contours are based on this rule.

3.2.2.2 Coarticulation Effect

When a syllabic sound is voiced, the following one is effected and vice versa. This effect on connected speech has been reported by Gandour, Potisuk and Dechongkit [26]. The study on tonal height reports that the anticipatory effect extends forward to about 75% of the duration of the following syllable and, similarly, the carry-over effect extends backward to about 50% of the duration of preceding syllable. This work uses the above duration during smoothing the F0 contour at the syllabic junction.

4. Issues in Speech Synthesis

After all necessary parameters for the synthesis are determined. This part will use these parameters to determine which sound units should be selected and how these units should be processed to synthesize a high natural synthetic speech. The parameters can be classified into two groups. First group is generated from the text analysis consisting of phonemic lists of a phrase. These lists are used to select the relevant units. Another one is generated from the prosody generation consisting of the duration and the F0 contours.

All are used in signal processing to improve the naturalness of synthetic speech. Before describing the detail of the synthesis techniques, the synthesis unit structure will be detailed.

4.1 Synthesis Scheme

In the synthesis work, there are various types of synthesis units used in concatenative speech synthesis such as words, syllables, demisyllables, phonemes, diphones, triphones etc. Each type has different advantages depending on the purpose of each system. In this work, demisyllable is selected because it has a reasonable number of sound units and acceptable quality. Although, its sound quality at syllable boundary is not quite natural as real speech but this problem can probably be improved by signal processing as being presented in this system.

4.1.1 Demisyllable

Demisyllable is the unit being the initial and final halves of a syllable. On the idea that a speech waveform is constructed by splicing the syllabic segment. A syllabic waveform is created from the proper initial and final demisyllable unit. Both units are segmented from a syllable at the stable vowel part. In general, Thai syllable has a structure of "C(C)VC" [1], therefore, the syllable is segmented into two portions, "C (C)V" and "VC". The initial unit contains a single consonant or double consonants linking with a partial vowel. The final unit contains a partial vowel linking with a final consonant and, also, tonal characteristic. Figure 4 shows an example of a demisyllable unit.

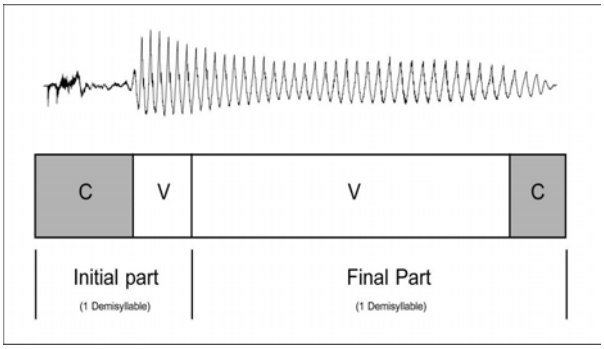


Figure 4. Demisyllable unit structure

4.1.2 Demisyllable inventory structure

In the previous section, we have already explained what the demisyllable-based concatenative synthesis is and how it works. Therefore, in this section we will discuss about how to list all units in the inventory and how many units are necessary in this system to construct the entire Thai syllables.

A Thai syllable sound can generally be characterized by four elements: initial consonant, vowel, final consonant and tone. Considering the Thai syllable structure, there are 4 patterns; CV (ปา [pa;0], รี [ri;1]), CCV (ปลา [pla;0], ครู [khru;0]), CVC (ปาด [pa;t1], กาก [ka;k1]), CCVC (ปราบ [pra;t1], กวาด [kwa;t1]). The first letter "C" represents an initial consonant while "CC" denotes a consonant cluster. The vowel of a syllable is represented by the letter "V" and the final consonant represented by the last letter "C". The Arabic numerals (0-4) represents the tone.

- Initial consonant: There are 44 consonantal letters in Thai. These letter represent 21 phonemes, grouped by traditional Thai grammarians into 3 classes; the high class, the middle class and the low class. These classes are very important in determining the tone of a syllable [6]. All of them can occur in the initial position such as กาด [ka;t], บน [bon], โยน [jo;n], etc. Some phonemes can be clustered to produce two different types of sound together such as ปล- /pl/, คว- /khw/, ตร- /tr/, etc. Our system also includes some phonemes for producing some loan words namely, consonant clusters and final consonants such as ดร- /dr/, ฟร- /fr/, บร- /br/, -ฟ /f/, -ล /l/, -ส /s/, etc. The list of Thai consonants is shown in Table 1 and 2.
- Vowel: There are 2 types of vowel; monophthong and diphthong. Monophthong can be classified into 2 groups; 9 long vowel sounds such as –า /a;/, ู /-u;/, แ- /x;/ and 9 short vowel sounds such as –ะ /a/, ุ /u/, แ-ะ /x;/, etc. Diphthong consists of 6 vowel sounds; 3 long vowel sounds such as เ-ีย /i;a/, เ-ือ /v;a/, -ัว /u;a/ and 3 short vowel sounds such as เ-ียะ /ia/, เ-ือะ /va/, -ัวะ /ua/ [27]. The list of Thai vowels is shown in Table 3.
- Final consonant: The final consonant consists of 9 phonemes, including open syllable such as ราด [ra;t1], กับ [kap1], จะ [ca?1], ยาว [ja;w0], etc. and 4 phonemes come from foreign language such as ball [bol], half [haf]. The reason of adding some borrowed phonemes in our inventory is that pronouncing a loan word close to the native pronunciation is preferable.
- Tone: There are 5 levels in Thai phonology represented in our inventory with the Arabic numerals (0-4); low-level [่ , 1], mid-level [no tone marker, 0], high-level [้ , 3], falling [๊ , 2] and rising [๋ , 4]. The syllable structure and the initial consonantal phoneme are very important to assign a tone for each syllable.

In this work, the demisyllable-based inventory for the initial part is constructed by creating all combination of 38 initial consonantal phonemes and 9 monophthongs (only short vowels sounds), resulting 342 units. The reason why only the short monophthongs are selected is the characteristics of short and long vowels are the same at the beginning. For final part, there are 1,163 units divided into 2 sets; 804 Thai phonemes and 359 phonemes of loan words. As a result, the total number of our inventory is 1,505 units. The combination is shown in table 4.

4.1.3 Construction phase

After a sound inventory has been designed, the next step is to select the speaker. A female speaker is selected based on the result of the listening test. All test sentences are read by 5 female speakers at a natural speaking speed and recorded directly to the computer. The speaker whose voice is natural and correctly pronounced in Thai standard is finally selected.

In our early work, the speaker pronounced a set of meaningless syllable sounds or logatoms and the recording was done in the office room environment, using a high quality microphone. The result was unsatisfactory because the synthesized speech was over stressed. Moreover, audible discontinuities occurred at the concatenated boundary. In the current work, we improve our synthesized speech by recording the target syllable with a frame sentence rather than syllable unit to solve the over stress problem and recording in the studio to get the clear voice. The female speaker reads each sentence 3 times for recording an initial part and a final part. From the recorded files, select the best one for manual segmentation. To segment a syllable into the initial and the final parts we have to find the segmenting point by listening and using the spectrogram to cut at a zero crossing point.

This process is an important step that has an effect on the quality of synthetic speech. However, such defects can be corrected by digital signal processing technique.

4.2 Speech Signal Processing

Applying signal-processing techniques to the synthesis part is a way to improve the naturalness. We apply the signal processing techniques (1) to normalize the amplitude of sound units to the same scale, (2) to capture the prosody parameter from the prosody generator into the synthetic speech waveform (this function makes change to pitch contour (F0 contour) and duration of speech signal.), (3) to smooth the discontinuity at the vicinity of concatenation. The considered discontinuities are the pitch variation, spectral mismatch and amplitude abruption.

4.2.1 Amplitude normalization

In a concatenation of sound units, the problem about abrupt change of amplitude at the concatenating point between the units. This problem occurs because each unit comes different syllable. For this reason, each unit has different amplitude. This defect makes synthetic speech sound fluctuated. To decrease this defect, this system normalizes the amplitude of all demisyllable-base inventory units to be the same standard.

First, the standard amplitude of each vowel is calculated. The frame sentence "เธอบอกให้พูด...ไปเรื่อยๆ" (/thq:?0 b@:k1 haj2 phu:t2 …. paj0 rv:aj2 rv:aj2/) with different target vowels in between is selected to pronounce and record for measuring the reference amplitude of the corresponding vowel. Then, each sentence is multiplied by a ratio that makes the mean of amplitude of the syllable that preceding the reference vowel, in this case is "พูด" (/phu:t2/), to be the same. Now, all reference vowel sounds present their own amplitude characteristic. These adjusted reference vowel sounds are used in calculating the amplitude reference of each vowel. Finally, these references are used in multiplying all demisyllable-based inventory units to have the same amplitude at the junction.

4.2.2 Prosodic modification

To modify prosody, this system uses Time-Domain Pitch Synchronous Overlap-Add (TD-PSOLA) [28] technique, a widespread technique which modifies pitch and duration of synthetic speech. This technique is based on dividing speech into subframes that partially overlapped on each other and each subframe is synchronized using pitch. To alter prosody, these subframes are slid to the preferred positions to modify prosody.

This technique is divided into three steps. First step, each unit in the entire inventory is marked at pitches in voiced part and, for unvoiced part, at virtual pitches every 4 millisecond as shown in Figure 5. All pitch markers are also labeled with its type whether voiced or unvoiced. Second step, the new pitch contour (F0 contour) is calculated. Actually, this new pitch contour was prepared from prosodic generating part. Next, the new pitch contour will be mapped with the old one using linear interpolation as shown in figure 6. This interpolation will specify which old pitch should be mapped to the new one so there are some old pitch missing or duplicating depending on this modification be to decrease or to increase duration. Final step, window function will process at each marked pitch on the original inventory units as shown in figure 7. Each window function is two-pitchwidth hanning window. By changing the displacement between marked pitch and overlapping some part of window at the edge, the prosody-modified synthetic speech is generated.

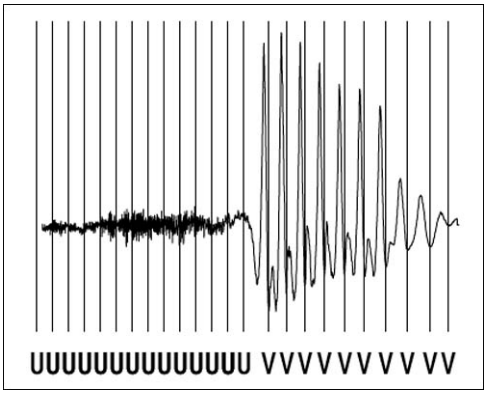


Figure 5. Pitch marking and labeling

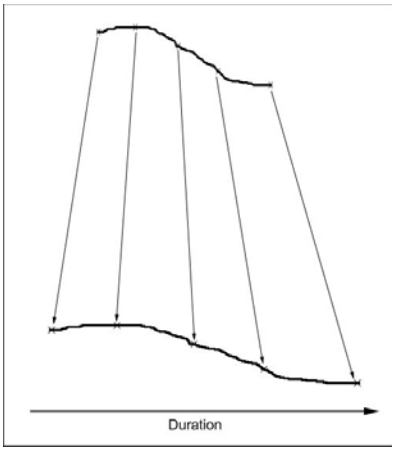


Figure 6 Duration Scaling.

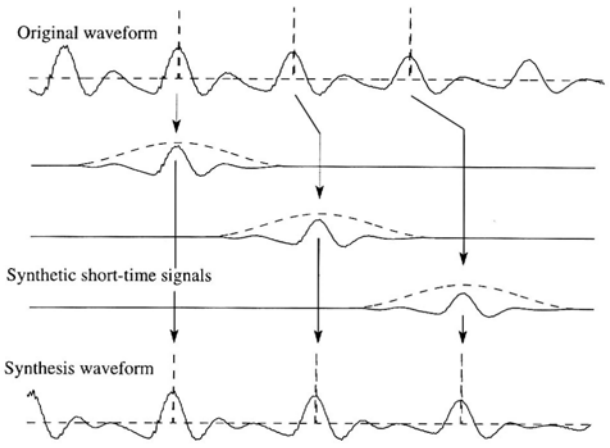


Figure 7. TD-PSOLA

4.2.3 Concatenated boundary smoothing

Due to the inventory unit used in this system is demisyllabic unit, so there appears some quality problems that happen at the intrasyllabic and intersyllabic concatenated point. These problems are a discontinuity of pitch, amplitude and, especially, spectrum. For the pitch discontinuity, it was solved in the prosodic modification part and, also, the amplitude discontinuity was already solved in the amplitude normalization part. The discontinuity of pitch and amplitude can be solved in time domain while the spectrum discontinuity must be solved in frequency domain. To solve the spectrum discontinuity, the speech signal is transformed to some representation in frequency domain. In this system, the Line Spectrum Pairs (LSP) [29] representation derived from LPC coefficients is selected.

The advantages of this representation are; (1)its parameters correspond to speech fomants, work like formant coding, (2) this representation is stable on interpolation. Figure 8 shown an example of a relation between formants computed by using LPC and their corresponding LSP parameters.

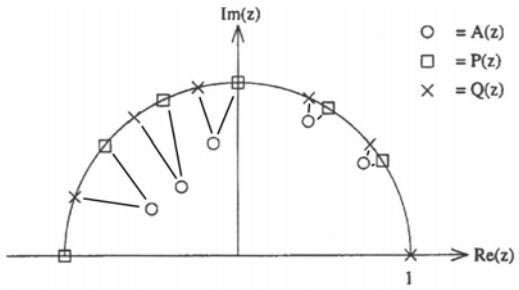


Figure 8. Example of relation between formants (circles) and its LSP parameters (squares and crosses)

This smoothing method is applied to this system to smooth the synthetic speech at the connecting points. To implement this method, first, several subframes at the junction are removed from the preceding demisyllable and the following one. Then the LSP parameter of subframes at the edges are computed as the reference parameters. The linear interpolation between these reference parameters is calculated to replace the removed subframes. This method is shown in figure 9.

4.2.4 Cross-syllable coarticulation modeling

When more than one syllable are connected together, there are some cross-syllable coarticualtion between adjacent syllables. These phenomena make the natural speech distinct from the synthetic speech. The effects of these phenomena are classified into two types. The first type is a prosody alteration, which is effected by the adjacent syllables. This effect was computed in the prosodic modification part. Another effect is a waveform interaction with neighbor syllables. Since the natural speech is the continuous speech but not the syllabic speech. Some syllabic speech signal is continuously transformed to the adjacent one. To improve the quality of the synthetic speech, this system includes these effects in synthesis part.

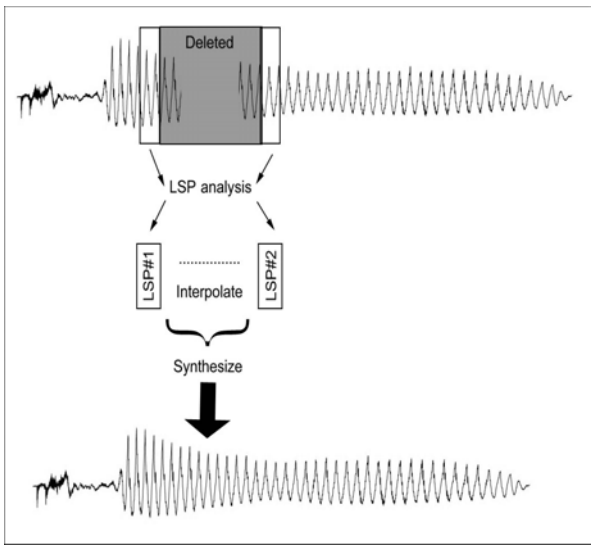


Figure 9. LSP Smoothing method

There are two connection types derived by investigating the natural speech. The first type is a simple touching. This occurs when the initial consonant of preceding syllable or final consonant of next one is unvoiced. Another one is an assimilated connection. This occurs when both the initial consonant of preceding syllable and final consonant of next one are voiced. In the implementation, this system uses the digital signal processing technique, LSP smoothing, as described above to simulate these connecting. The example is shown in figure 10. Figure (10a) shows the simple touching type and the assimilated connecting type is shown in figure (10b).

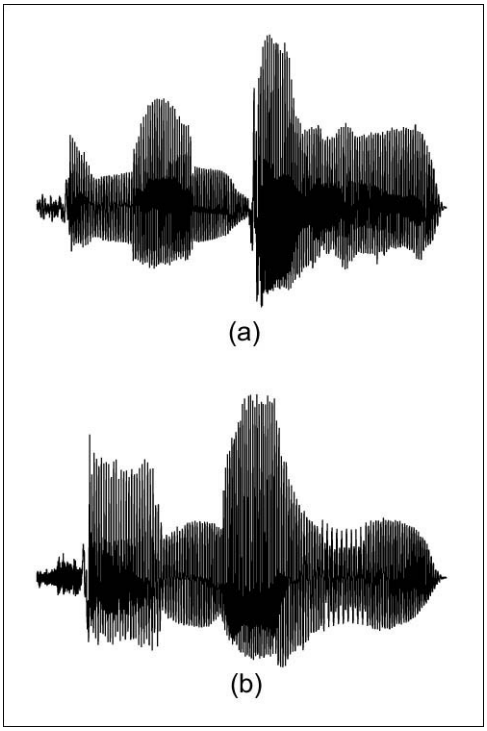


Figure 10. Syllable connections:
(a) a simple touching
(b) an assimilated connection

5. Future work

After implementing the system with the above approach and evaluating the synthetic speech, we found that the quality was acceptable. However, the improvement of the naturalness of the speech is suggested.

In the aspect of text analysis, there are two major points: the input text segmentation and the grapheme-to-phoneme conversion that need improvement. The improvement in grapheme-to-phoneme conversion will boost the system in the way to synthesize the correctly pronounced speech but not the naturalness. The most important problem is how to handle the homograph disambiguation. The homograph is that a word has more than one possible phonological representation such as the word 'เพลา' can be transcribed into '/phe:0-la:0/' or '/phlaw3/' depending on its context. The input text segmentaion includes the tasks of sentence extraction and phrase break determination. This task has an affect on the naturalness of synthetic speech because its result implies the determination of the pause position that has the major role in the prosody generation. The break positions are so ambiguous that even the native Thai people also can not judge decisively whether they are actual break. The rule-based approach has been introduced for the task. An advantage of rule based approach is that it is easy to developed. But it also has a drawback in handling the problem that has several interacting factors and high degree in ambiguity like the prosody parameter prediction task. The statistical or corpus based method is another alternative. In the future, we plan to apply the corpus-based approach to both the text analysis and prosody generation.

In the aspect of speech synthesis, other acoustic inventory stucture such as diphone, triphone, syllable and the non-uniform unit, etc. is in our consideration. Also the automatic unit selection algorithm which works well in other language systems is studied to replace the manual speech segmentation in inventory construction phase. Furthermore, other advance topics such as the voice transformation and more sophisticated synthesizer are also our interesting topics.

To pursue these future works, it is evident that the large-scale prosody-labeled speech corpus be indispensible for us. Our next step is to design and to develop the speech corpus which be labeled with the complete information.

Acknowledgement

We would like to thank Ms.Tanakorn Wiboon, an intern student from Department of Computer Engineering, Kasertsart University, for her contribution in the implementation of grapheme-to-phoneme conversion module.

References

[1] Luksaneeyanawin, S. et., al., 1992. A Thai text-to-speech system, Proceeding of 4th NECTEC Conference, pp.65-78 (in Thai).

[2] Taisertavattanakul, S. and Kanawaree, W., 1995, Thai speech synthesizer. Unpublished senior project report, Department of Computer Engineering, Chulalongkorn University (in Thai).

[3] Kiat-arpakul, R., Fakcharoenphol, J. and Keretho, S., 1995, A combined phoneme-based and demisyllable approach for Thai speech synthesis, Proceedings of the 2nd Symposium on Natural Language Processing SNLP'95, pp 361-369.

[4] Luksaneeyanawin, S., 1995, Tone transformation, Proceedings of the 2nd Symposium on Natural Language Processing SNLP'95, pp. 345-360.

[5] Hansakunbuntheung, C., Leelarasmee, E., 1999, Thai syllabic speech synthesis based on line spectrum pair, 22nd Thailand Electrical Engineering Conference, pp.521-524 (in Thai).

[6] Luksaneeyanawin, S., 1993. Speech computing and speech technology in Thailand, Proceedings of the Symposium on Natural Language Processing in Thailand, pp.276-321.

[7] Danvivathana, N., 1987, The Thai writing system, Forum Phoneticum 39, Helmut Buske Verlag Hamburg.

[8] Thavaranon, K., 1978, Spacing in Thai Writing, M.A.Thesis Department of Thai Chulalongkorn University (in Thai).

[9] Mittrapiyanuruk, P. and Sornlertlamvanich, V., 2000, The automatic Thai sentence extraction, Proceeding of 4th Symposium on Natural Language Processing (SNLP'2000).

[10] Sornlertlamvanich, V., 1993, Word segmentation for Thai in machine translation system, Machine Translation, NECTEC pp 556-561 (in Thai).

[11] Thavaranon, K., 1978, Spacing in Thai writing, Master Thesis, Department of Thai, Chulalongkorn University (in Thai).
[12] Sileverman, K., 1987, The Structure and Processing of Fundamental Frequency Contours, Ph.D. Thesis, University of Cambridge.
[13] Karn, H., 1996, Design and evaluation of a phonological phrase parser for Spanish text-to-speech, Proceedings of the Fourth International Conference on Spoken Language Processing, Vol. 3, pp. 1696-1699.
[14] Dutoit, T., 1997, Introduction to text-to-speech synthesis, Kluwer Academic Publishers.
[15] Campbell, W.N. and Isard, S.D., 1991, Segment durations in a syllable frame, Journal of Phonetics, Vol. 19, pp37-47.
[16] Luangthongkum, T., 1977, Rhythm in standard Thai, Unpublished Ph.D. Thesis, Univerity of Edinburge.
[17] Klatt, D.H., 1987, Review of text to speech synthesis conversion for English, Journal of Acoustic Society America, Vol 82, pp.737-793.
[18] t'Hart, J., and Cohen, A., 1973, Intonation by rule: a perceptual quest., Journal of Phonetics, 1:309:327.
[19] t'Hart, J., and Collier, R., 1975, Integrating different levels of intonation analysis, Journal of Phonetics, 3:235-255.
[20] Pierrehumbert, J. B., 1980, The phonology and phonetics of English intonation., PhD Thesis, Published by University of Edinburgh.
[21] Cooper, W.E. and Sorensen,J.M., 1981, Fundamental Frequency in Sentence Production., Springer-Verlag., 1981.
[22] Liberman, M. and Pierrehumbert, J., 1984, Intonational invariance under changes in pitch range and length., In Aronoff, M. and Oehrle, R T., editors, Language Sound Structure., MIT Press.
[23] Fujisaki, H. and Kawai, H., 1988, Realization of linguistic information in the voice fundamental frequency contour of the spoken Japanese, In International Conference on Speech and Signal Processing. IEEE.
[24] Taylor, P.A., 1992, A Phonetic Model of English Intonation, PhD Thesis, Edinburgh.
[25] Luksaneeyanawin, S., 1983, Intanation in Thai, Unpublished PhD Thesis, University of Edinburgh.
[26] Gandour, J. T., Potisuk, S., and Dechongkit, S., 1994, Tonal Coarticulation in Thai, Journal of Phonetics, vol 22, pp.477-492.
[27] Khanitthanan, W., 1990, Phasa lae Phasasart, Thammasat University Press. (in Thai)
[28] Charpentier, F. and Moulines, E., 1989, Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones, European Conference on Speech Communication and Technology, vol. I, pp. 013-019.
[29] Klejin W.B. and Paliwal, K.K., 1995, Speech coding and synthesis, Elsevier Science.

Virach Sornlertlamvanich is the acting director of Information Research and Development Division of the National Electronics and Computer Technology (NECTEC) of Thailand since 1992. He received the B.Eng. and M.Eng. degrees from Kyoto University, in 1984 and 1986, respectively. From 1988 to 1992, he joined NEC Corporation and involved in the Multi-lingual Machine Translation Project supported by MITI. He received the D.Eng. degree from Tokyo Institute of Technology in 1998. His research interests are natural language processing, lexical acquisition and information retrieval.

Pradit Mittrapiyanuruk received bachelor degree in electrical engineering from King Mongkut's University of Technology Thonburi (KMUTT) in 1994 and master degree in electrical engineering Chulalongkorn University in 1996. Then he joined NECTEC in August 1996. He had involved in the projects i.e. Integrated Receiver&Decoder (IRD), Text Retrieval Database, and speech synthesis. Currently, he mainly works for the NECTEC's Thai text-to-speech synthesis project. His research interests are speech synthesis, speech recognition and multimedia signal processing.

Chatchawarn Hansakunbuntheung received his Bachelor degree and Master degree in electrical engineering from Chulalongkorn University in 1998 and 2000 respectively. He has joined the NECTEC at the Software and Language Engineering Laboratory (SLL) since 2000. He started his work in the research and development group of the Thai Text-to-Speech project since. At the present, he is involved in Thai Text-to-speech project and Thai speech corpus project. His research interests are speech technology and Natural Language processing.

Virongrong Tesprasit has joined NECTEC in April 1996 after receiving her BA.(Linguistics) degree from Thammasat University. She had joined both Royal Institute Dictionary Development Network Project and Development of Thai Corpus Base Project. At present, she researches on Thai Text-to-Speech synthesis Project. Her research interests are Phonetics and Speech Technology.

Table 1. Phonetic symbol of Thai consonant

Thai Letter	Phonetic Symbol	
	Initial	Final (including open syllable)
ก	/k/	/k/
ข, ฃ, ค, ฅ, ฆ	/kh/	
ง	/ng/	/ng/
จ	/c/	
ฉ, ช, ฌ	/ch/	
ซ, ศ, ษ, ส	/s/	
ญ, ย	/j/	/j/
ฎ, ด	/d/	/t/
ฏ, ต	/t/	
ฐ, ฑ, ฒ, ถ, ท, ธ	/th/	
ณ, น	/n/	/n/
บ	/b/	/p/
ป	/p/	
พ, ภ, ผ	/ph/	
ฟ, ฝ	/f/	
ม	/m/	/m/
ร	/r/	
ล, ฬ	/l/	
ว	/w/	/w/
ห, ฮ	/h/	-
อ	/?/	-

Table 2. Consonant Cluster

Thai Letter	Phonetic Symbol of Consonant Cluster		English Letter	Phonetic Symbol of Consonant Cluster	
	Initial	Final		Initial	Final
ปร-	/pr/	-	br-	/br/	-
ปล-	/pl/	-	bl-	/bl/	-
ตร-	/tr/	-	fr-	/fr/	-
กร-	/kr/	-	fl-	/fl/	-
กล-	/kl/	-	dr-	/dr/	-
กว-	/kw/	-	f-	-	/f/
พร-, ผร-	/phr/	-	l-	-	/l/
พล-, ผล-	/phl/	-	s-	-	/s/
ทร-	/thr/	-	ch-	-	/ch/
คร-, ขร-	/khr/	-	-	-	-
คล-, ขล-	/khl/	-	-	-	-
คว-	/khw/	-	-	-	-

Table 3. Phonetic Symbol of Thai Vowel

Monophthong				Diphthong				Vowel Letter			
Short Vowel		Long Vowel		Short Vowel		Long Vowel		Short Vowel		Long Vowel	
-ะ	/a/	-า	/a;/	เ-ียะ	/ia/	เ-ีย	/i;a/	ำ	/am/	-	-
ิ-	/i/	ี-	/i;/	เ-ือะ	/va/	เ-ือ	/v;a/	ไ-, ใ-	/aj/	-	-
ึ-	/v/	ื-	/v;/	ั-วะ	/ua/	ั-ว	/u;a/	เ-า	/aw/	-	-
ุ-	/u/	ู-	/u;/								
เ-ะ	/e/	เ-	/e;/								
แ-ะ	/x/	แ-	/x;/								
โ-ะ	/o/	โ-	/o;/								
เ-าะ	/@/	-อ	/@;/								
เ-อะ	/#/	เ-อ	/#;/								

Table 4. Combination of Demisyllable based Inventory for Final Part

Vowel	Final Consonant		
	Dead Syllable (3)	Live Syllable (5)	Open Syllable (1)
Monophthong			
Short vowels (9)	Mid, Low, Falling, High	Mid, Low, Falling, High	Mid, Low, Falling, High
Long vowels (9)	Mid, Low, Falling, High	Mid, Low, Falling, High, Rising	Mid, Low, Falling, High, Rising
Diphthong			
Short vowels (3)	-	-	Low, Falling, High
Long vowels (3)	Low, Falling, High	Mid, Low, Falling, High, Rising	Mid, Low, Falling, High, Rising

Toward an Enhancement of Textual Database Retrieval By using NLP Techniques*

Asanee Kawtrakul[1], Frederic Andres[2], Kinji Ono[2],
Chaiwat Ketsuwan[1], Nattakan Pengphon[1],
ak@beethoven.cpe.ku.ac.th , {andres,ono}@rd.nacsis.ac.jp

(1) NAiST(1), Computer Engineering Dept, Kasetsart University, Bangkok, Thailand
(2) NACSIS(2), Center of Excellence of the Ministry of Education, Tokyo, Japan

ABSTRACT : Improvements in hardware, communication technology and database have led to the explosion of multimedia information repositories. In order to provide the quality of information retrieval and the quality of services, it is necessary to consider both retrieval techniques and database architecture.

This paper presents the project named VLSHDS-Very Large Scale Hypermedia Delivery System. The quality of textual information search is enhanced by using NLP techniques. The quality of service over a large-scale network is provided by using AHYDS-Active HYpermedia Delivery System-framework.

KEY WORDS : Information Retrieval, Textual Database Retrieval, Multi-level indexing, Document Classification, Very Large Scale Hypermedia Delivery System, Natural Language Processing

บทคัดย่อ : การพัฒนาเทคโนโลยีฮาร์ดแวร์ เทคโนโลยีสื่อสาร และฐานข้อมูลได้นำไปสู่การเติบโตอย่างรวดเร็วของการจัดเก็บข้อมูลแบบหลายสื่อ เพื่อให้การบริการข้อมูลและการสืบค้นข้อมูลมีคุณภาพและประสิทธิภาพ เราจำเป็นต้องพิจารณาทั้งทางด้านเทคนิคการสืบค้นข้อมูลและสถาปัตยกรรมฐานข้อมูล

บทความฉบับนี้นำเสนอผลงานวิจัยภายใต้โครงการที่ชื่อว่า ระบบจัดส่งข้อมูลหลายสื่อขนาดใหญ่ (VLSHDS) ด้วยเทคนิคการประมวลภาษาธรรมชาติในระดับคำ และระดับวลี สามารถยกระดับคุณภาพของการสืบค้นข้อมูล ด้วยเทคโนโลยีของระบบจัดส่งข้อมูลหลายสื่อแบบแอคตีฟ สามารถเพิ่มคุณภาพการให้บริการข้อมูล

คำสำคัญ : การสืบค้นข้อสนเทศ การสืบค้นข้อมูลเอกสาร การสร้างดัชนีหลายระดับ การแยกประเภทเอกสาร ระบบจัดส่งข้อมูลหลายสื่อขนาดใหญ่

- This Project has been granted by Kasetsart University Research and Development Institute (KURDI), Kasetsart University, Thailand and National Center for Science Information Systems (NACSIS), Center of Excellence of the Ministry of Education, JAPAN and National Electronics and Computer Technology Center (NECTEC).
- This article is a reprint of the article appeared in the Proceedings of NECTEC Annual Conference 2000 : ECTI Technologies for New Economies, June 2000, pp. 280-290. This paper wins a best paper award in category of "the most impact to Thai society".

1. Introduction

Improvements in hardware, communication technology and database engines had led to the expansion of challenging interactive multimedia applications and services. Typical examples of applications include on-line news, digital libraries and web-based information involving multi-dimension multimedia document repositories. These systems combine various media content with hyperlink structures for user query or navigation. Most of them store contents inside the database systems supporting extenders in order to add application data types with their access methods. Moreover, there is no vertical integration between application plug-ins and the database kernel itself. This limitation is an underlying reason for further improvements [6,8,16,17,18]. AHYDS-The Active HYpermedia Delivery System is one of a new wave of database kernels [4,7,15] that facilitates the access to multimedia documents according to the user's requirement and application's features over a wide spectrum of networks and media [1].

The VLSHDS-Very Large Scale Hypermedia Delivery System is the project between NACSIS and NAiST [2,10] which is aimed to integrate both the quality of data management service and the quality of textual information retrieval. The VLSHDS platform is , then, based on AHYSD which provides a framework for open data delivery service, communication service, query execution service and supervision service. The quality of full text retrieval services has been enhanced in both precision and recall by integrating NLP techniques for document and query processing.
Section 2 gives an overview of the VLSHDS. The implementation of document processing, query processing and retrieving processing are described in section 3, 4 and 5 respectively. Section 6 gives the conclusion and briefs the next step of the project.

2. An Overview of the Very Large Scale Hypermedia Delivery Systems

The key architectural components in the VLSHDS platform used as textual database platform is shown in Figure 1. The system consists of a client/server three tiers architecture. At Client side, queries are sent to the server by using the AHYDS communication support [11]. At the server side, there are three main components: Document Processing, Query Processing and Retrieving Processing. The Document Processing based on the Extended Binary Graph (EBG) structure provides multilevel indices and document category as document representation. The Query Processing provides query expansion based on query guide. The Retrieval Processing computes the similarity between queries and documents and returns a set of retrieved documents with the similarity scores.

3. The Role of NLP in Document Processing

To enhance the performance of full text retrieval service, good representation of each document should be provided, i.e., multi-level indices and document category. Multi-level indices will increase the retrieval recall without the degradation of the system precision and document category will be used for pruning irrelevant document or increase precision while decreasing the searching time.

The primary problem in computing multi-level indices and category as document representation is a linguistic problem. The problems frequently found, especially in Thai documents, are lexical unit extraction including unknown word, phrase variation, loan words, acronym, synonym and definite anaphora.

Accordingly, to be more successful, NLP components, i.e., morphological analysis and shallow parsing should be integrated with statistical based indexing and categorizing

3.1 The Architecture of NLP based Document Processing

Figure 2 shows the overview of Thai document processing. There are two main steps: multilevel indexing and document categorizing. Each document will be represented as

$$D_i = \langle I_p, I_t, I_c, C_i \rangle$$

Where I_p, I_t, I_c are the set of indices in phrase, single term and conceptual level, respectively

Ci is the category of an document i-th

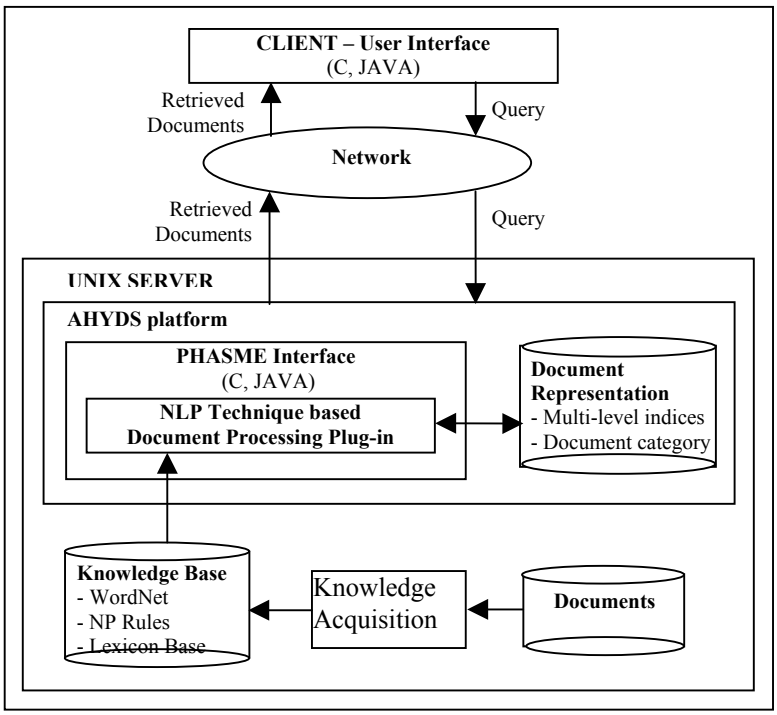


***Figure 1**: The Architecture of the VLSHDS Platform for Textual Document Retrieval*

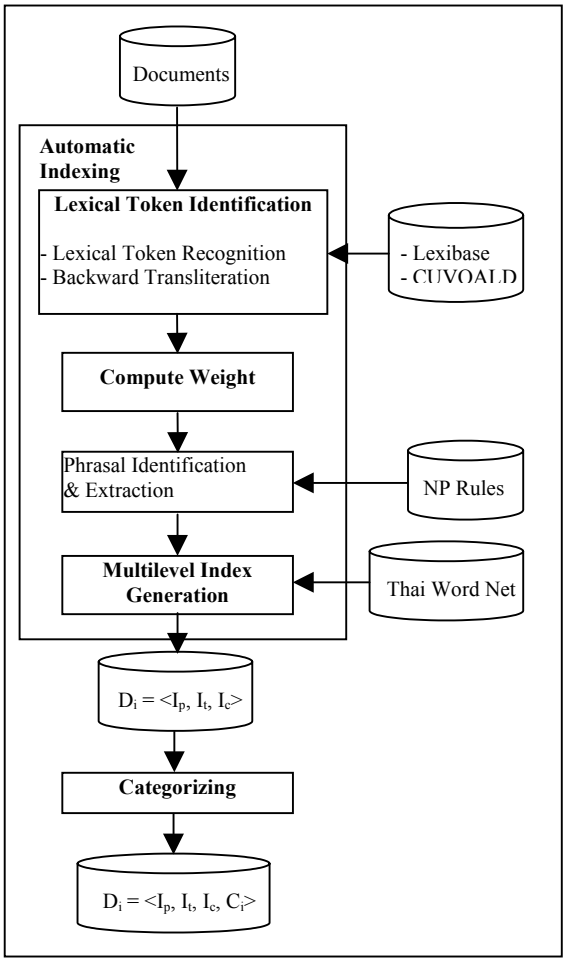


***Figure 2**: Overview of Thai document processing*

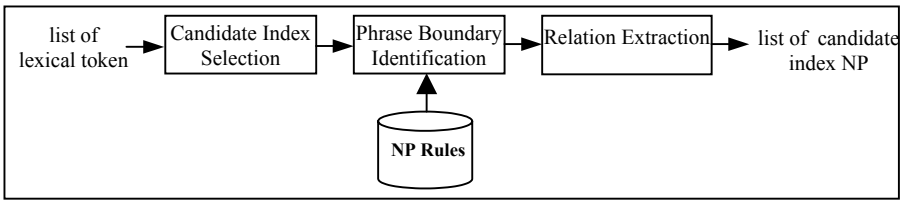


***Figure 3**: Phrase Identification and Relation Extraction*

NP <- cn + cn + (cn)	กล้วยไม้ป่า
NP <- cn + {cn, pn}	ประเทศไทย
NP <- cn + v + (cn)	วิทยุกระจายเสียง
NP <- cn + (cn) + mod	น้ำปลาหวาน
NP <- cn + prep + cn	คุณค่าทางอาหาร
NP <- cn + v + v	ค่าใช้จ่าย
NP <- cn + num + cl	สัตว์สองเท้า
NP <- cn + NOM	อุตสาหกรรมการกลั่นสุรา
NOM <- prefix + vp	การกลั่นสุรา

***Figure 4**: Noun phrase rules*

3.2 Automatic Multilevel Indexing

As shown in Figure 2, automatic multilevel indexing consists of three modules: lexical token identification, phrase identification with relation extraction, and multilevel index generation. Each module accesses different linguistic knowledge bases stored inside the EBG data structure.

3.2.1 Lexical Token Identification

The role of lexical token identification is to determine word boundaries, and to recognize unknown words, such as proper names and loan words, i.e., technical terms in Thai orthography. Based on [12], lexical tokens are segmented and tagged with part of speech. Based on [13], loan word will be solved.

3.2.2 Phrase Identification and Relation Extraction

In order to compute multilevel indices, phrase identification and relation extraction are needed. The relation between terms in the phrase will be extracted in order to define indices in single term level and conceptual level. There are two kinds of relations: head-modifier and non-head-modifier (or compound noun). If the relation is head-modifier, the head part will be the single term-level index while the modifier will not be used as index. If the relation is compound noun, there is no single term-level index. The conceptual level indices, then, will be produced from the head of phrase or compound noun by accessing Thai wordnet.

The problems of this step are that (1) how to identify phrase without deep parsing for the whole text, (2) how to distinguish between noun phrase and sentence which may have the same pattern and (3) how to extract the relation between terms in phrase.

To solve the problems mentioned above, the algorithm of phrase identification and relation extraction consists of three main steps (see Figure 3):

The first step is the candidate index selection which is based on Salton's weighting formula [9], then, the second step is the phrase boundary identification by using statistical based NLP technique. This step will find phrase boundary for a set of candidate indices(provided from the first step) by using NP rules (see Figure 4)

At this step, we can describe as 4-tuple:

$$\{T, N, R, NP\}$$

where

T is the set of candidate terms which have weight $w_i > \theta$ (θ is a threshold)

N is the set of non-terminal

R is the set of rules in the grammar

NP is the starting symbol

The third step is the relation extraction. After the boundary of phrase(s) is identified, we need to compute the relation between a set of words in the phrase in order to find whether that NP is head-modifier NP or compound. If the frequency of each word of candidate NP has the same frequency (see Figure 5) then

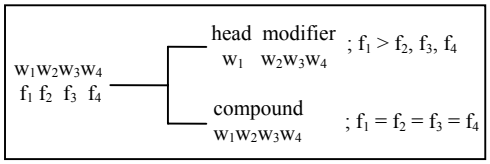


***Figure 5**: Relation Extraction*

relation of that NP is compound noun, otherwise relation of that NP is head-modifier pair. Head is the term(s) with highest frequency. Modifier is the term(s) with lower frequency.

The detail of the algorithm is given in Annex 1.

3.2.3 Multilevel index generation

At this step, each document D_i is summarized and represented in the EBG data structure as a vector of numeric weights, i.e.:

$$D_i =< I_{p_i}, I_{t_i}, I_{c_i} >$$

where

$$I_{p_i} =< W_{p1_i}, W_{p2_i}, ..., W_{p3_i} >$$
$$I_{t_i} =< W_{t1_i}, W_{t2_i}, ..., W_{t3_i} >$$
$$I_{c_i} =< W_{c1_i}, W_{c2_i}, ..., W_{c3_i} >$$

Phrasal level indices (I_p) consist of set of the phrases extracted by using noun phrase rules. Single term level indices (I_t) are the head of each index token in the phrasal level. Conceptual level indices (I_c) are the semantic concepts of each single term level index, given in Lexibase [14]. For example, the document concerns about มะนาว (lemon), may keep "มะนาวไข่" (A kind of lemon) as phrasal level and keep "มะนาว" (Lemon) as single term level and "พืช" (Plant) as conceptual level.

Figure 6 shows the process of Multilevel Index Generation consisting of phrase level, single term level and conceptual level for each document.

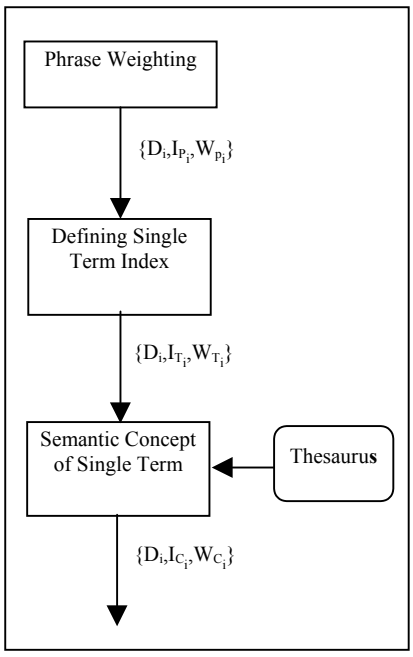


***Figure 6**: Multilevel Index Generation*

In each level of index we use Salton's Weight normalization [15] as shown below is used for computing weights.

$$V_k^{\,m} = \frac{tf_k^{\,m} \log \frac{N_d}{n_k}}{\sum_{j=1}^{l} tf_j^{\,m} \log \frac{N_d}{n_j}}$$

tf_k^m = Number of index terms k in document m
n_k = Number of documents that contain term k
N_d = Number of documents in the collection
l = Number of index terms

The parallel processing of the document is providing by the AHYDS engine using the EBG data structure. Each level of index of each document is computed dependently but independently from other document.

More details of the algorithm of multilevel index generation is given in Annex 2.

Figure 7 shows the comparison between multilevel indexing and traditional indexing system

Using multilevel indexing, "egg" would not be retrieved, while, in traditional IR, it will be retrieved which degrade the performance of the system.

3.2.4 Document Classification

Even though multi-level indices can cover a very wide range of document retrieval without degradation of system performance, document clustering for pruning irrelevant documents is still

$$D_i =< I_{p_i}, I_{t_i}, I_{c_i}, C_i >$$

necessary in order to increase precision and decrease searching time.

Text categorization or document clustering consists of two parts: a prototype learning process to provide prototypes for each cluster of documents and a clustering process, which compute the similarity between input document and prototype (see Figure 8).

Finally, document will be represented as

where

$$I_{p_i} =< W_{p1_i}, W_{p2_i}, ..., W_{pt_i} >$$
$$I_{t_i} =< W_{t1_i}, W_{t2_i}, ..., W_{tt_i} >$$
$$I_{c_i} =< W_{c1_i}, W_{c2_i}, ..., W_{ct_i} >$$
$$C_i =< W_{cat1_i}, W_{cat2_i}, ..., W_{catn_i} >$$

The algorithm of document clustering is summarized in Annex 3.

4. Query processing

In order to obtain those documents, which have the best match with a given query, we also need a "query guide". Query Guide is applied by using the cluster hypothesis and query expansions. Our method apply Word-Net for reconstructing query by adding more general term/concept. For example

Query = "น้ำดอกไม้." (proper name: the name of mango) and its general term (from Word-Net) is "มะม่วง" (mango). After expansion, the new query is "มะม่วง-น้ำดอกไม้" (The phrase contains of mango and its specified name).

5. Retrieval Processing

The following retrieval process is implemented for enhancing the performance of the system (see figure 9):

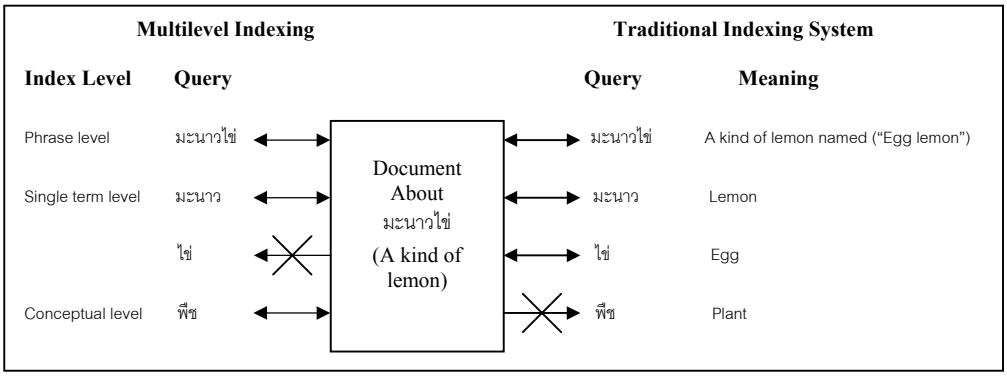


***Figure 7**: Example of how Multi-Level Indexing can enhance performance*

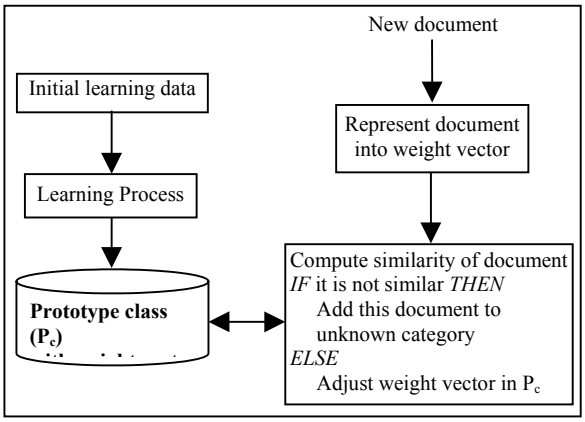


***Figure 8**: Text Categorization process*

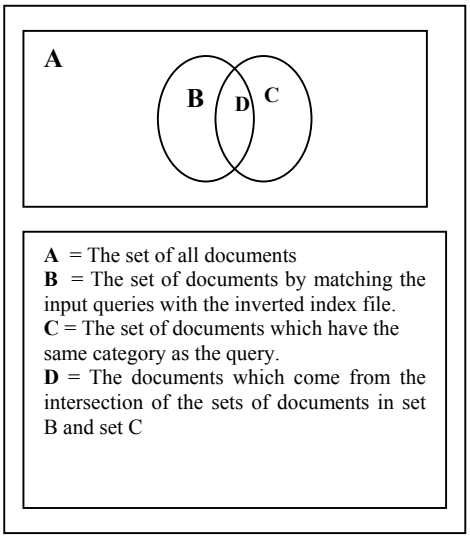


***Figure 9**. The Retrieved Documents*

1. Compute a candidate set of documents by matching the input queries with the inverted index file.
2. Select a candidate set of documents which have the same category as the query.
3. Calculate the similarity between the queries of the documents which come from the intersection of the sets of documents in 1 and 2.
4. Return a set of retrieved document with the similarity scores.

6. Conclusion and Future Work

Figure 10 shows the difference between a set of index with applying and without applying NLP techniques.

At the current state, knowledge acquisition is processed manually by linguists. Next step it will be provided by semi-automatically. The domain of documents is limits in computer area. However, it will be extended to cover agriculture and general news area.

References

[1] Andres F., "Active Hypermedia Delivery System and PHASEMA Information Engine" in Proc. First International Symposium on Advanced Informatics, Tokyo, Japan, 2000.

[2] Andres F., Kawtrakul A., Ono K. and al., "Development of Thai Document Processing System based on AHYDS by Network Collaboration", in Proc. 5th international Workshop of Academic Information Networks on Systems(WAINS), Bangkok, Thailand, December 1998.

[3] Andres F., and Ono K. "The Active Hypermedia Delivery System", in Proceedings of ICDE98, Orlando, USA, February 1998.

[4] Boncz, P.A. and Kerstern, M.L. "Monet: An Impressionist Sketch of an Advanced Database System" In Proc. IEEE BITWIT Workshop, San Sebastian (Spain), July 1995.

[5] E. Chaniak, "Statistical Language Learning", MIT Press, 1993.

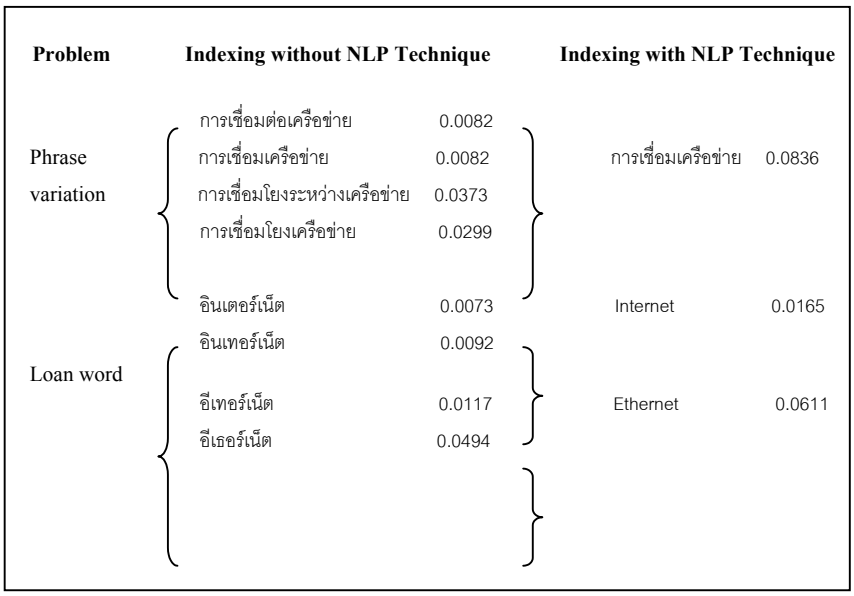


***Figure 10**. Examples of Applying NLP Technique to solving phrase variation and loan word*

[6] Geppert, A. Dittrich, K.R. "Constructing the Next 100 Database Management Systems: Like the Handyman or Like the Engineer ?" in SIGMOD RECORD Vol.23, No 1, March 1994.

[7] Geppers A., Scherrer S., and Dittrich K.R. "Kids: Construction of Database Management Systems based on Reuse", Technical report 97.01, Institutfur Informatik, University of Zurich, Switzerland, 1997.

[8] Grosky W.I. "Managing Multimedia Information in Database System" in Communication of the ACM December 1997, Vol 40., No 12, page 73-80.

[9] G. Salton, "Automatic Text Processing. The Transformation, Analysis, and Retrieval of Information by Computer", Singapore: Addison-Wesley Publishing Company, 1989.

[10] Kawtrakul A., Andres F., et.al.,. "A Prototype of Globalize Digital libraries: The VLSDHS Architecture for Thai Document processing." 1999. (on the process of submission)

[11] Kawtrakul A., Andres F., Ono K. and al., "The Implementation of VLSHDS Project for Thai Document Retrieval" in Proc. First International Symposium on Advanced Informatics, Tokyo, Japan, 2000.

[12] Kawtrakul A., et.al., "Automatic Thai Unknown Word Recognition", In Proceedings of the Natural Language Processing Pacific Rim Symposium, Phuket, pp.341-346, 1997.

[13] Kawtrakul A., et.al., "Backward Transliteration for Thai Document Retrieval", In Proceedings of The 1998 IEEE Asia-Pacific Conference on Circuits and Systems, Chiangmai, pp. 563-566, 1998.

[14] Kawtrakul A., et.al., "A Lexibase Model for Writing Production Assistant System" In Proceedings of the 2nd Symposium on Natural Language Processing, Bangkok, pp. 226-236, 1995.

[15] Seshadri P., Livny M., and Ramakrishnan R. "The Case for Enhanced Abstract Data Types" In Proceedings of 23rd VLDB Conference, Athens, Greece, 1997, pages 56-65.

[16] Subrahmanian V.S. "Principles of Multimedia Database Systems", Morgan Kaufmann, 1997.

[17] Teeuw W.B., Rich C., Scholl M.H. and Blaken H.M. "An Evaluation of Physical Disk I/Os for Complex Object Processing" in Proc. IDCE, Vienna, Austria, 1993, pp 363-372.

[18] Valduriez P., Khoshafian S., and Copeland G. "Implementations techniques of Complex Objects" in Proc. Of the International Conference of VLDB, Kyoto, Japan, 1986, pp 101-110.

Biography

Assoc. Prof. Asanee Kawtrakul,Ph.D., researcher, received bachelor degree (honor) and master degree in electrical engineering from Kasetsart University in 1976 and 1986, respectively. She received Ph.D in Information Engineering from Nagoya University in 1991. She has been the lecturer for Faculty of Engineering since 1983. She had been the project leader in several projects in the fields of Natural Language Processing, Text Retrieval Database, Speech Synthesis, Database Management System, and Geographical Information System.

Annex 1

Algorithm Phrase Identification and Relation Extraction

Input: a list of lexical token $w_1, w_2, \ldots, w_n$
with set of POS tag information $T_i = \{t_1, t_2, t_3, \ldots, t_m\}$,
frequency f_i and weight W_i for each word

Output: set of candidate index NPs with head-modifier relation
or compound relation

Candidate Index Selection:
Selecting candidate index by selecting term which have weight $w_i > \theta$
(θ is an index threshold)

Phrase boundary identification:
FOR each candidate term *DO*
Apply NP rule to find boundary
IF can not apply rule directly
Consider weight of adjacent term w_{adj} *THEN*
IF adjacent term has weight $w_{adj} > \phi$
(ϕ is a boundary threshold) *THEN*
Extend boundary to this term
ELSE
IF this adjacent term in the preference list
Extend boundary to this term *THEN*

Relation Extraction:
FOR each candidate index phrase *DO*
To find internal relation we consider term frequency
in each phrase
IF the frequency of each word of candidate NP
has the same frequency *THEN*
Relation of this NP is *compound noun*
ELSE
Relation of this NP is *head-modifier pair* :
Head is the term(s) with highest frequency.
Modifier is the term(s) with lower frequency.

Annex 2

Algorithm Multilevel Index Generation

Input: 1. A list of lexicon token provided by Lexicon Token Identification and Extraction process w1, $w_{2,....}$, w_j with its frequency f_{wi} and weight w_{wi}.
2. A list of candidate Phrasal indices with head-modifier relation or compound relation.

Output: Index weight vector as document representation

$$D_i = \{I_{p_i}, I_{t_i}, I_{c_i}\}$$

Where

$$I_{p_j} = \{w_{p_1}, w_{p_2}, ..., w_{p_j}\}$$
$$I_{t_j} = \{w_{t_1}, w_{t_2}, ..., w_{t_j}\}$$
$$I_{c_j} = \{w_{c_1}, w_{c_2}, ..., w_{c_j}\}$$

Phrasal Level Indexing:

FOR each candidate Index NP *DO*
Recompute Phrase weights in whole documents
IF phrase weight > θ (θ *is* index threshold)
Keep sorted Phrase index token *THEN*

Single Term Level Indexing:

FOR each candidate phrase index NP *OR* each single term *DO*
IF tokens of candidate phrasal index
Extract the head of the token *THEN*
Recompute weight
ELSE
FOR each lexicon token that not appear in phrasal level *DO*
Recompute weight
Keep sorted Single term indices

Conceptual Level Index:

FOR each single term index *DO*
Find Semantic Concept of each single terms
Recompute weight
Keep Sorted Conceptual Level Index

Annex 3

Algorithm Document Clustering

Input: single term indices and phrase indices with their frequencies

Output: Document representation in $<I_{p_i}, I_{t_i}, I_{c_i}, C_i>$

Learning Process:

Define prototype class e.g. Computer, Agriculture, News etc.

FOR each documents *DO*

Compute weight vector of single term/phrasal indices by using

$$W_x{}^m = \frac{tf_x{}^m \log \frac{N_d}{n_x}}{\sum_{j=1}^{l} tf_j{}^m \log \frac{N_d}{n_j}}$$

W_x^m = Weight Vector of phrase indices x in document mth
x = k mean Single term index
x = p mean Phrasal index
tf_x^m = {0 if $f_x^m = 0$, $\log(f_x^m) + 1$ otherwise}
n_k = Number of documents that contain term *k*
N_d = Number of documents in the collection
l = Number of index terms

FOR each prototype class *DO*

Compute weight vector of single term indices by using Rocchio's algorithm is used [3, 5]:

$$W_{ck} = \begin{cases} 0 & \text{Otherwise} \\ W'_{ck} & \text{If } W'_{ck} > 0 \end{cases}$$

$$W'_{cx} = \beta \frac{1}{|R_c|} \sum_{i \in R_c} W_x^m - \gamma \frac{1}{\left|\overline{R}_c\right|} \sum_{i \in \overline{R}_c} W_x^m$$

W_{cx} = weight of term k in the prototype P_c for class.
x = k = Single Term index
x = p = Phrasal index
W_x^m = weight of term k indexing for each document
x = k = Single Term index
x = p = Phrasal index
R_c = set of training documents belonging to class c
$\overline{R}_c$ = set of documents not belonging to class c
(Note: $\beta = 16$, $\gamma = 4$ [Buckley et al., 1994])

Classification:

FOR new document input *DO*

Compute weight of phrase index and weight of single term index by using formula:

$$W_x{}^m = \frac{tf_x{}^m \log \frac{N_d}{n_x}}{\sum_{j=1}^{l} tf_j{}^m \log \frac{N_d}{n_j}}$$

W_x^m = Weight Vector of phrase indices x in document mth
x = k mean Single term index
x = p mean Phrasal index
tf_x^m = {0 if $f_x^m = 0$, $\log(f_x^m) + 1$ otherwise}
n_k = Number of documents that contain term *k*
N_d = Number of documents in the collection
l = Number of index terms

Compare weight with each Prototype Class by using the dot product formula

$$S(D_i, C) = \alpha \frac{\sum_{k=1}^{t} (W_p^m * W_{cp})}{\sqrt{\sum_{k=1}^{t} (W_p^m)^2 * \sum_{k=1}^{t} (W_{cp})^2}} + \beta \frac{\sum_{x=1}^{t} (W_k^m * W_{ck})}{\sqrt{\sum_{x=1}^{t} (W_k^m)^2 * \sum_{x=1}^{t} (W_{ck})^2}}$$

W_{ck} = Weight of Single term *k* in the Prototype P_c for class
W_k^m = Weight of Single term *k* in document mth
W_{cp} = Weight of Phrase *p* in Prototype P_c for class
W_i^m = Weight of Phrase *p* in document mth
$\alpha = [0, 1]$
$\beta = [0, 1]$
$\alpha + \beta = 1$

FOR each C, *DO*

IF $S(D_i, C) > \theta_t$ *THEN*

Store document into Prototype P_c and adjust weight in Prototype Class.

การพัฒนาช่องว่างสำหรับพิมพ์ข้อมูลที่มีระบบการตรวจสอบความถูกต้อง
Development of data entry box with data validity system

รศ. วรชัย ตั้งวรพงศ์ชัย ถวัลย์ สุขทะเล
หน่วยรังสีรักษา ภาควิชารังสีวิทยา คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น
รุจชัย อึ้งอารุณยะวี
ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยขอนแก่น

บทคัดย่อ-- การพัฒนาช่องว่างเพื่อให้ผู้ใช้เติมข้อมูลลงในแบบบันทึกข้อมูล มีความจำเป็นต้องพัฒนาระบบการตรวจสอบความถูกต้องของข้อมูลที่พิมพ์ลงในช่องว่าง วิธีการตรวจสอบที่นำเสนอในบทความนี้ อาศัยจากแนวคิดที่ว่า ช่องว่างที่พัฒนาขึ้นจะมีพฤติกรรมในการทำงานเหมือนช่องว่างทั่วไป เมื่อผู้ใช้พิมพ์คำหรือข้อความไม่ตรงกับที่ผู้พัฒนากำหนดไว้ ระบบก็จะเตือนพร้อมกับแสดงรายการของข้อความที่ผู้ใช้ควรจะเติม กรณีที่ผู้ใช้พิมพ์ไม่สมบูรณ์ ระบบก็จะจัดหาข้อความที่สมบูรณ์และมีความเป็นไปได้มากที่สุดเติมลงในช่องว่างให้ ด้วยวิธีการนี้จะไม่ส่งผลกระทบต่อการทำงานของผู้ใช้ แต่จะมั่นใจได้ว่าระบบจะได้ข้อมูลที่น่าเชื่อถือที่สุด

โมดิฟายด์อิเลกโทรดสำหรับวิเคราะห์ปริมาณกลูตาเมท
Modified Electrode for the Determination of Glutamate

นางพรพิมล ศรีทองคำ : สถาบันพัฒนาและฝึกอบรมโรงงานต้นแบบ
รศ. บุษยา บุนนาค, รศ.ดร. มรกต ตันติเจริญ : คณะทรัพยากรชีวภาพและเทคโนโลยี
ผศ.ดร. โสพส สุวรรณยืน : ภาควิชาวิศวกรรมเคมี คณะวิศวกรรมศาสตร์
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี

บทคัดย่อ--โมดิฟายด์อิเลกโทรดสำหรับวิเคราะห์ปริมาณกลูตาเมทสร้างขึ้นจากส่วนผสมของผงคาร์บอน,อนุภาคขนาดเล็กของรูทีเนียม, เอนไซม์ glutamate dehydrogenase, (GLDH) และ โคเอนไซม์ NAD+ สภาวะที่เหมาะสมต่อการเตรียมอิเลกโทรดคือการผสมคาร์บอนที่มีโลหะรูทีเนียมร้อยละ 1 (น้ำหนัก/น้ำหนัก) ปริมาณ 12.5 มิลลิกรัมกับสารละลายเอนไซม์ GLDH และ NAD+ ในฟอสเฟตบัฟเฟอร์น้ำหนัก 5 และ 3.8 มิลลิกรัมตามลำดับ จากนั้นสร้างฟิล์มบางของ poly(1,3-diaminobenzene)บนผิวหน้าของอิเลกโทรดดังกล่าวด้วยวิธีอิเลกโตรเคมิคอลโพลีเมอไรเซชัน ศักย์ไฟฟ้าที่ใช้ในการวัดกลูตาเมทโดยโมดิฟายด์คาร์บอนเพสอิเลกโทรดคือ 400 มิลลิโวลต์ เทียบกับขั้วอ้างอิง Ag/AgCl อิเลกโทรดมีกิจกรรมคงเหลือ 80% หลังการใช้งาน 100 ครั้ง เวลาในการตอบสนองประมาณ 2-3 นาที พิสัยเชิงเส้นของการวัดกลูตาเมทอยู่ในช่วง 0.01-0.9 มิลลิโมลาร์

การตรวจวินิจฉัยสัญญาณไฟฟ้ากล้ามเนื้อลิ้นและคอขณะกลืน ด้วยอิเล็กโตรดชนิดปิดผิวหนัง

Surface Electromyography in Dysphagia

วิทูร ลีลามานิตย์, แอนดรูว์ ซีการ์ , อลัน กีเตอร์
สถาบันวิศวกรรมชีวการแพทย์ มหาวิทยาลัยสงขลานครินทร์
เครือข่ายศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ

บทคัดย่อ-- วัตถุประสงค์ของโครงการวิจัยปีที่ 1 ระยะที่ 2 คือศึกษาลักษณะจำเพาะของ surface electromyo- graphy (sEMG) ของกล้ามเนื้อลิ้นและคอ (tongue and thyrohyoid muscle) ขณะอาสาสมัครกลืนน้ำลายและอาหารชนิดต่างๆ วิธีวิจัย ทำการบันทึก sEMG ของกล้ามเนื้อลิ้นและคอในอาสาสมัครจำนวน 61 คน ขณะอาสาสมัครกลืนน้ำลาย น้ำ 5 มิลลิลิตร เยลลี่ 5 มิลลิลิตร ขนมปัง (biscuit) ขนาด 5 มิลลิลิตร อย่างละ 3 ครั้ง และทำการบันทึก sEMG ของกล้ามเนื้อลิ้นและคอในอาสาสมัครอีก 2 คน คนละ 3 ครั้งห่างกันครั้งละ 1 สัปดาห์ ขณะอาสาสมัครกลืนน้ำลาย น้ำ 5 มิลลิลิตร และ 10 มิลลิลิตร อย่างละ 6 ครั้ง เพื่อทดสอบ reproducibility ของวิธีตรวจวัด และ intrasubject and intersubject variation ทำการประมวลผลของ sEMG ทั้งหมดด้วย algorithm ที่ใช้ในระยะแรกของโครงการวิจัย แล้ววิเคราะห์ลักษณะจำเพาะของ sEMG ด้วยวิธี 1.หาค่ารากที่สองของผลคูณค่าเฉลี่ยพื้นที่ใต้ curve (SRMAUC) ของ sEMG กล้ามเนื้อลิ้นและคอขณะกลืน 2.หาผลรวม vector (CV) ของ sEMG กล้ามเนื้อลิ้นและคอขณะกลืน 3.นำค่าSRMAUC และ CV ของ sEMG ในข้อ 1 และ 2 มาหาความสัมพันธ์ ผลการวิจัยพบว่า 1.SRMAUC ของ sEMG กล้ามเนื้อลิ้นและคอขณะกลืนขนมปังจะแตกต่างอย่างมีนัยสำคัญจากการกลืนน้ำลาย กลืนน้ำ และเยลลี่ ($p<.001$) และ ค่า SRMAUC ขณะกลืน น้ำลายจะแตกต่างอย่างมีนัยสำคัญจากการกลืนเยลลี่ ($p<.05$) ส่วนค่า SRMAUC ขณะกลืนน้ำจะไม่แตกต่างจากการกลืนน้ำลายและเยลลี่ ($p>.05$) 2.ค่า CV ของ sEMG กล้ามเนื้อลิ้นและคอขณะกลืนขนมปังแตกต่างอย่างมีนัยสำคัญจากการกลืนน้ำลาย น้ำ และเยลลี่ ($p<.001$) และค่า CV ขณะกลืนเยลลี่จะแตกต่างอย่างมีนัยสำคัญจากการกลืนน้ำลายและกลืนน้ำ ($p<.05$) ส่วนค่า CV ขณะกลืนน้ำจะไม่แตกต่างจากกลืนน้ำลาย ($p>.05$) 3.ค่า SRMAUC และค่า CV ของ sEMG มีความสัมพันธ์ในรูปสมการยกกำลังโดยมี $R2=0.83$ สรุป การหาค่า SRMAUC และค่า CV ของ sEMG กล้ามเนื้อลิ้นและคอขณะกลืนด้วยวิธีดังกล่าวข้างต้นสามารถใช้ศึกษาลักษณะจำเพาะของ sEMG กล้ามเนื้อลิ้นและคอขณะกลืนในอาสาสมัครปกติได้

Technical Journal
60 Vol II, No.7, March-June, 2000

พีโซอิเลกตริกคริสตัลไบโอเซนเซอร์สำหรับวิเคราะห์ยาปราบศัตรูพืชกลุ่มออร์กาโนฟอสฟอรัส : การโมดิฟายด์ผิวหน้าอิเลกโทรดด้วย poly (1,3-diaminobenzene) โดยเทคนิค Electrochemical Polymerisation

Piezoelectric Crystal Biosensor for the Detection of Organophosphorus Pesticides : Modification of Electrode Surface with Poly (1,3-diaminobenzene) by Electrochemical Polymerisation Technique

นางพรพิมล ศรีทองคำ, นายธวัชชัย สุวรรณ : สถาบันพัฒนาและฝึกอบรมโรงงานต้นแบบ
รศ.ดร. มรกต ตันติเจริญ : คณะทรัพยากรชีวภาพและเทคโนโลยี
ดร.กฤษณพงศ์ กีรติกร : สายวิชาเทคโนโลยีวัสดุ คณะพลังงานและวัสดุ
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี

บทคัดย่อ-- งานวิจัยนี้กล่าวถึงวิธีการโมดิฟายด์ขั้วอิเลกโทรดทองของพีโซอิเลกตริกคริสตัลด้วยโพลีเมอร์ poly (1,3-diamonobenzene) โดยเทคนิค electrochemical poly-merisation อิเลกโทรดที่ผ่านการโมดิฟายด์จะถูกตรึงด้วยเอนไซม์ acetylcholinesterase และนำไปทดสอบการตอบสนองต่อสารประกอบออร์กาโนฟอสฟอรัสในตัวอย่างน้ำประเภทต่างๆ ไบโอเซนเซอร์ที่สร้างขึ้นจากสภาวะที่เหมาะสมสามารถตอบสนองต่อไดคลอวอสด้วยความสัมพันธ์อย่างเชิงเส้นในช่วงความเข้มข้น 0.05-2.0 พีพีเอ็ม และพบว่าไบโอเซนเซอร์สามารถตอบสนองต่อคาร์บาเมทได้ แต่ไม่พบการตอบสนองต่อสารประกอบกลุ่มอื่นเช่น ออร์กาโนคลอรีน ยาปราบวัชพืชหรือไอออน นอกจากนี้ได้ศึกษาถึงการฟื้นฟูกิจกรรมของไบโอเซนเซอร์โดยสารประกอบ 2-PAM เพื่อศึกษาความเป็นไปได้ในการนำไบโอเซนเซอร์กลับมาใช้ใหม่

VLSI Implementation of a Symmetric Cipher Using Cellular Automata

Banlue Srisuchinwong, Thitiporn Lertrusdachakul,
Orapin Watcharawetsaringkan and Kittipong Meesawat
Department of Electrical Engineering, Sirindhorn International Institute of Technology
Thammasat University, Rangsit Campus, Pathumthani, 12121, Thailand

บทคัดย่อ-- บทความนี้เสนอการสร้างวงจรรวมขนาดใหญ่สำหรับวงจรป้องกันการดักฟัง (Cipher) แบบสมมาตร โดยใช้ เซลลูล่า ออโตเมต้า แบบ non-autonomous และแบบ autonomous โดยการใช้ข้อมูลทุก ๆ 16 บิท ผ่านเข้าไปใน เซลลูล่า ออโตเมต้าแบบ non-autonomous ข้อมูลสามารถไหลทางเดียวได้โดยการใช้ involutions การใช้เซลลูล่า ออโตเมต้าแบบ autonomous จะเปลี่ยนรหัสกุญแจขนาด 96 บิทไปตลอดเวลาในขณะที่ข้อมูลผ่านเข้ามา โครงการนี้ ได้ออกแบบวงจรโดยใช้หลักการ "บนลงสู่ล่าง (Top-Down Design)" ใน 3 ขั้นตอนคือ behavioural level (C language and logic simulations) และ structural level (transistors and spice simulations) สำหรับขั้นตอนที่ 3 คือ physical level (layout) นั้น ยังมิได้นำเสนอในบทความนี้ การออกแบบ "ล่างขึ้นบน (Bottom-Up Design)" ได้นำมาใช้ด้วย ทำให้ได้ข้อสรุปว่า วงจรสามารถทำงานได้ถึง 21 เมกกะเฮิรซ์ โดยมีความเร็ว 336 เมกกะบิทต่อวินาที คุณสมบัติที่สำคัญของเซลลูล่า ออโตเมต้า ได้แก่ ความเรียบง่าย ความเป็นมอดุลาร์ และการติดต่อสื่อสารภายในที่ใช้ระยะทางเพียงสั้นๆกับเซลข้างเคียง คุณสมบัติเหล่านี้ล้วนเหมาะสมอย่างยิ่งสำหรับการสร้างวงจรไฟฟ้ารวมขนาดใหญ่มาก

A High-Speed Multiplier-Free Realization of IIR Filter Using ROM'S

Thanyapat Sakunkonchak and Sawasd Tantaratana
Department of Electrical Engineering, Sirindhorn International Institute of Technology
Thammasat University, Rangsit Campus, Pathumthani, 12121, Thailand
E-mail: thong@siit.tu.ac.th, sawasd@siit.tu.ac.th

บทคัดย่อ-- ในบทความนี้ ผู้เขียนได้นำเสนอไอไออาร์ฟิลเตอร์ความเร็วสูงแบบไร้ตัวคูณโดยการใช้รอม เพื่อเก็บผลคูณกับสัมประสิทธิ์การคูณ ควบคู่ไปกับสัญญาณความถี่สูงและการทำไปป์ไลน์ ด้วยการปรับค่าของตัวแปรบางตัวทำให้โครงสร้างที่นำเสนอนี้ได้ค่าของฮาร์ดแวร์และความเร็วที่แตกต่างกัน ดังตัวอย่างซึ่งได้ทำการเปรียบเทียบโครงสร้างที่นำเสนอกับ Distributed Arithmetic ผลลัพธ์ที่ได้แสดงให้เห็นว่าถ้าเลือกตัวแปรที่เหมาะสมแล้ว โครงสร้างที่นำเสนอนี้จะให้ความเร็วที่สูงกว่าและใช้ฮาร์ดแวร์น้อยกว่าเมื่อเทียบกับ DA realization

วิธีการปกปิดข้อผิดพลาดในการถอดรหัสวิดีโอภาพระบบ H.261 โดยการใช้วิธีการประมาณค่าของเวคเตอร์การเคลื่อนที่

An Error Concealment Method for the H.261Video Decoding Using Estimation of the Motion Vectors

*เทอดศักดิ์ ธนกิจประภา *นงนุช สุขตั้งมั่น **ไกรสิน ส่งวัฒนา **อิทธิชัย อรุณศรีแสงไชย
*นักศึกษาปริญญาโท คณะวิศวกรรมศาสตร์ **อาจารย์ คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหาร ลาดกระบัง

ABSTRACT - The detection and concealment of error in the H.261 standard video stream is important in any video conferencing session as error in the bitstream can affect not only its corresponding picture element but also other neighbouring elements in the same frame and also other frames that references the errornous element. The error can propagate until the starting of a new Group of Block (GOB) if it is not detected.

This paper presents a method for concealment of error in macroblock upon detection of error while decoding the H.261 bitstream. Temporal error concealment using motion vector from current and previous frames is applied. An array of macroblocks for concealment which formulated from motion vectors in current and previous P-frames by average, interpolate and extrapolate. Then, the best macroblock to conceal the error macroblock is selected.

KEYWORDS – GOB , Macroblock

Application of Inclusion Scheduling to Resource Estimation in Architectural Synthesis With Imprecise Specification

Chantana Chantrapornchai
Dept. of Mathematics Silpakorn University

Sissades Tongsima
HPCC, NECTEC

ABSTRACT – In this paper, we apply *inclusion scheduling* to estimate resource bounds in architectural synthesis for VLSI systems. The inclusion scheduling algorithm takes an application which may consist of imprecise information and generates a good schedule on average. The framework for resource estimation considers the design goal, and first creates the initial bound. Then inclusion scheduling is used as a tool to adjust the bound while considering imprecise information.

KEY WORDS – architectural synthesis, resource estimation, scheduling, allocation, imprecise information

ระบบควบคุมสภาพแวดล้อมระยะไกลสำหรับเครือข่าย

อภิเนตร อูนากูล, มีลาภ โสขุมา, เอกชัย วิวรรธนาภิรักษ์
ห้องปฏิบัติการ Embedded System (ESL) ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ABSTRACT – The Remote Environmental Control System is a system that can be used to remotely control electronic equipment through various communication media such as telephone, modem, remote control, and computer through serial port. This system is intended to control and monitor the status of the network equipment located in the rural areas unattended by technical personnel in the Government Information Network (GINET) project. This paper presents the design of the system in three levels, the physical design, the RS485 protocol design, and the software design. The software design uses the object-oriented design technique and design pattern to reduce implementation complexity which lead to less development time, easier validation, and ease of maintenance.

KEY WORDS -- Embedded System, Home Networking, Information Appliance

Applying ATM Network Technology for IMT-2000
(การประยุกต์ใช้เทคโนโลยีเครือข่ายเอทีเอ็มสำหรับ IMT-2000)

ผศ. ดร. สินชัย กมลภิวงศ์ ธัธชัย เองฉ้วน อัมพิกา จันทรภักดี
สุธน แซ่หว่อง มัลลิกา อุณหวิวรรธณ์
Department of Computer Engineering, Faculty of Engineering,
Prince of Songkla University, Hatyai, Songkla, Thailand 90112

Abstract-- In this paper, we present an investigation of ATM network technologies for IMT-2000. Our study work will focus on the ATM deployment in co-operate with IMT-2000. We first show some limitations of AAL-1 when it is used for carrying low bit rate voice. In contrast, when assessing the use of AAL-2 for carrying low bit rate voice, AAL2 offers a number of advantages when compared with AAL-1. We briefly reviewed network architecture issues related to internetworking model. We have addressed some challenging issues which may be concerned for future research topics.

Keywords: ATM, IMT-2000, Mobile, Wireless, Internetworking

A High-Speed Multiplier-Free Realization of IIR Filter Using ROM'S

Thanyapat Sakunkonchak and Sawasd Tantaratana
Department of Electrical Engineering, Sirindhorn International Institute of Technology
Thammasat University, Rangsit Campus, Pathumthani, 12121, Thailand
E-mail: thong@siit.tu.ac.th, sawasd@siit.tu.ac.th

ABSTRACT – In this paper, we propose a high-speed multiplier-free realization using ROM's to store the results of coefficients scalings in combination with higher signal rate and pipelined operations. By varying some parameters, the proposed structure provides various combinations of hardware and clock speed (or throughput). An example is given comparing the proposed realization with the distributed arithmetic (DA) realization. Results show that with proper choices of the parameters the proposed structure achieves a faster processing speed with less hardware, as compared to the DA realization.

KEY WORDS – IIR filter, multiplier-free realization, pipelined realization.

การเข้ารหัสเสียงพูดแบบอะแด็พทีฟดิฟเฟอร์เรนเชียลพัลส์โค้ดมอดูเลชั่น โดยใช้เทคนิคออโต้คอรีเลชั่นคู่

Speech Coding by Adaptive Differential Pulse Code Modulation using Dual Autocorrelation Technique

**วรการ วงศ์สายเชื้อ **ไกรสิน ส่งวัฒนา*
**นักศึกษาปริญญาโท คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง*

ABSTRACT – This paper presents speech coding by Adaptive Differential Pulse Code Modulation (ADPCM) by using the dual autocorrelation to predict the signal. Generally, the autocorrelation of adjacent speech sample is greater than the autocorrelation of several order time delayed samples. So, the prediction of present sample by using the one past and next sample is more effective than only using the several order time delayed past samples. This prediction method is used in ADPCM encoding and the result compares with the standard ADPCM encoding.

KEY WORDS – ADPCM , Dual autocorrelation

Wavelength Routing Switching using Birefringent Fiber By Coupled Polarization Modes

P.P. Yupapin
*Lightwave Technology Research Center, Department of Applied Physics, Faculty of Science King Mongkut's Institute of Technology Ladkrabang (**KMITL**), Bangkok 10520, Thailand Tel: 6627373000 ext. 6271, Fax: 3269981, E-mail :Yupapin.Preecha@kmitl.ac.th*

ABSTRACT-- This paper presents the study of an optical signal processing scheme known as a wavelength routing switching where the selected wavelength channel is routed by stretching a polarization maintaining fiber. The principle of the scheme is that the wavelength channel multiplexing signals are orthogonaly combined then propagated in a single mode polarization maintaining fiber. The desired wavelength channels could be controlled by stretching, i.e. coupling, the employed fiber length. Results obtained using two multiplexed wavelength channels of 670 and 632.8 nm sources have shown the measured crosstalk of -7 dB, where the signal to noise ratio of 14 dB was achieved.

Keywords-- Optical switching, Optical devices

วิธีการมอดูเลทและดีมอดูเลทสัญญาณดิจิตอลในย่านความถี่วิทยุผ่านเครือข่ายโทรทัศน์ชนิดใช้สาย นำสัญญาณแบบ BPSK/QPSK

BPSK/QPSK Radio Frequency Data Modulation and Demodulation for Cable Television Network

** ทิษณุ งามเธียรธนา **ไกรสิน ส่งวัฒนา*
**นักศึกษาปริญญาโท คณะวิศวกรรมศาสตร์ **อาจารย์ คณะวิศวกรรมศาสตร์*
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหาร ลาดกระบัง

ABSTRACT--This paper presents radio frequency modulator and demodulator for high speed data communication on Hybrid Fiber Coaxial cable television network. Modulation and demodulation uses quadrature phase shift keying, which have changes in carrier phase by multiple of 90 degree.

The modulator consists of : digital signal processor, carrier signal processor, signal multiplier and signal summation. The demodulator consists of : QPSK amplifier, clock generator, carrier recovery signal generator, signal multiplier and digital processor.

The resultant modem operates at 10 MHz carrier frequency with data rate 1.28 Mbps/S.

KEYWORDS – QPSK , Cable Modem

การวิเคราะห์ประสิทธิภาพการสื่อสารข้อมูลในโครงข่าย HFC ภายใต้สภาพแวดล้อมของสัญญาณรบกวนแบบอิมพัลส์บนเส้นทางกลับ

Performance Analysis of Data Communication in HFC Network Under Impulsive Noise Environment on the Return Paths

ชวลิต ชั้นไพบูลย์[*] กัญน์ สิทธิประเสริฐ[*] ทิษณุ งามเธียรธณา[*] ไกรสิน ส่งวัฒนา[**] อิทธิชัย อรุณศรีแสงไชย[**]

นักศึกษาปริญญาโท คณะวิศวกรรมศาสตร์[**] อาจารย์ คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ABSTRACT - This paper analyses the performances of data communication on the return path of the HFC Network, following the DOCSIS (Data Over Cable Service Interface Specifications) and IEEE 802.14 standards. Three methods of modulation; BPSK QPSK and 16 QAM are analyzed under impulsive noise environment with reference to class-A impulse noise model. The result indicates the relationship of each modulation methods in term of BER via CNR.

KEY WORDS – Hybrid Fiber Coaxial Network , Return Path , Impulsive Noise , Carrier to Noise Ratio

การใช้ภาพสแกน X-ray Film 2 มิติ เพื่อหาตำแหน่ง 3 มิติ ของแท่งแร่ที่ได้รับการรักษาแบบใส่แร่ภายใน

3D of Radioisotope source positioning in Radioisotope insertion Treatment by using an Imaging of X-ray Film scan

ถวัลย์ สุขทะเล, รศ.วรชัย ตั้งวรพงศ์ชัย

หน่วยรังสีรักษา ภาควิชารังสีวิทยา คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น

ABSTRACT – The position of radioisotope source in a patient who was treated by radioisotope insertion technique is very important for calculate a dose distribution. The Inputting from keyboard or digitizer is good practice for source positioning but not enough for source to organ relationship in a patient. Inputting source position on the computer screen together with x-ray imaging will be done easier and can be seen the source to organ relationship that very helpful in reducing of inputting error. An error may be rise from a measurement on the film by using ruler or the film slid while using the digitizer. The inputting on screen will produce an error less than 0.3 mm. (depend on the screen resolution). The x-ray imaging can be improved the quality by adjust a contrast, brightness, Zoom in/out and Pan the image on the screen. It can be show the relationship line between the antero-posterior image with respect to the lateral image as well.

KEY WORDS – radioisotope source position, radioisotope insertion.

การตรวจหาเชื้อมาเลเรียในภาพเซลล์เม็ดเลือดแดงโดยการเปรียบเทียบมัลติพีคของฮิสโตรแกรม

The Detection of Malaria Parasite in Red Blood Cells Image by Multi-Peak Histogram Comparison

กริช สมกันธา สัญญา คล่องในวัย* สมชัย เอี่ยวสานุรักษ์* บุญธีร์ เครือตราชู***

นักศึกษาบัณฑิตศึกษา อาจารย์ภาควิชาวิศวกรรมคอมพิวเตอร์***

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าคุณทหารลาดกระบัง

Abstract-- This research present the detection of malaria parasite in red blood cells image by using computer. In the detection of malaria parasite in red blood cells, the blood film must be dyed. Then scan the desired color. The malaria parasite is indicated by the color. The intense level of malaria parasite which it is different from red blood cells. So the research present the detection of malaria parasite in red blood cells image by Multi-Peak histogram comparison. The Multi-Peak histogram comparison can be create software computer for the detection of malaria parasite. In the experiment uses 30 image of red blood cells. From the experiment result, the Multi-Peak histogram comparison can be used to detect malaria parasite with 100 percent accuracy.

KEY WORDS – Medical Image ,Enhancement and Segmentation

เทคนิคการสกัดลักษณะเด่นโครงร่างแบบเวกเตอร์ของอักษรภาษาไทย โดยวิธีสุ่มแบบเป็นลำดับตามเส้นอักษร

วิทยากร แซ่มกัน, อัศนีย์ ก่อตระกูล

ห้องปฏิบัติการวิจัยสารสนเทศอัจฉริยะ ภาควิชาวิศวกรรมคอมพิวเตอร์ มหาวิทยาลัยเกษตรศาสตร์

ABSTRACT-- This paper presents a method for extract skeleton vector from Thai print characters base on continuous sampling on character lines. The major techniques for skeleton vector is calculate point's direction on character lines [1]. But the recognition is not good enough because it still keep noises to be the feature. We correct this problem with continuously sampling on character line with junctions and then delete sample points that is not essential (not represent the sharp curve or sharp angle of line) after that calculate the direction of vectors. Data structure support multi objects character.

KEY WORDS -- vector feature, sampling and skeleton

Toward an Enhancement of Textual Database Retrieval By using NLP Techniques

Asanee Kawtrakul[1], Frederic Andres[2], Kinji Ono[2],
Chaiwat Ketsuwan[1], Nattakan Pengphon[1],
ak@beethoven.cpe.ku.ac.th , {andres,ono}@rd.nacsis.ac.jp
(1) NAiST[(1),] Computer Engineering Dept, Kasetsart University, Bangkok, Thailand
(2) NACSIS[(2)], Center of Excellence of the Ministry of Education, Tokyo, Japan

ABSTRACT-- Improvements in hardware, communication technology and database have led to the explosion of multimedia information repositories. In order to provide the quality of information retrieval and the quality of services, it is necessary to consider both retrieval techniques and database architecture.

This paper presents the project named VLSHDS-Very Large Scale Hypermedia Delivery System. The quality of textual information search is enhanced by using NLP techniques. The quality of service over a large-scale network is provided by using AHYDS-Active HYpermedia Delivery System-framework.

การควบคุมแขนหุ่นยนต์ข้อต่อเดียวแบบอ่อนตัวด้วยการควบคุมขั้นสูง

Advanced Control of One-Link Flexible Robot Arms

ผศ.ดร. วัชรพงษ์ โชวิฑูรกิจ ดร. มานพ วงศ์สายสุวรรณ และ ดร. เดวิด บรรเจิดพงศ์ชัย
ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ABSTRACT -- This work deals with the application of advanced control techniques, namely, adaptive, robust and intelligent control, in the control of one-link flexible robot arms so as to force it to move to the desired positions and to reduce the vibration that occurs during the motion due to its flexible nature. The objectives are (i) to determine the feasibility, performance, advantages and disadvantages of various techniques mentioned above and (ii) to develop advanced control software

KEY WORDS – one-link flexible robot arm, adaptive control, robust control, intelligent control

การวิเคราะห์โครงสร้างทางกลและเสถียรภาพของหุ่นยนต์เดินสองขา

ฐิติศักดิ์ จันทร์พรหม, ไพศาล สุวรรณเทพ, ชิต เหล่าวัฒนา
ศูนย์ปฏิบัติการพัฒนาหุ่นยนต์ภาคสนาม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
91 ถ.ประชาอุทิศ แขวงบางมด เขตทุ่งครุ กรุงเทพ 10140
โทร 66(2)470-9335 โทรสาร 66(2)470-9339, E-Mail: s1400172@cc.kmutt.ac.th

ABSTRACT-- This paper describes our preliminary research in analyzing mechanical structure and its stability of a humanoid robot, to be designed and built at FIBO. Mobility and gaits of such a robot are governed by only two legs, leading to high complexity in dynamic control. We have thoroughly measured positions, velocities and accelerations of each joint of human legs in order to understand their profiles. We are in a process of designing geometry of the robot legs, based on such measured profiles. In addition, we have proposed the robot gaits with related analysis on kinematic and dynamic stability. Finally, a planar two degree of freedom inverted pendulum (PTIP) has been built as a testbed to implement our controller for balancing one leg. Simulation results and discussion are included herein.

KEY WORDS -- Humanoid robot, Stability, Gait

การพิมพ์ภาพกราฟฟิกส์โดยใช้เลเซอร์พลอตเตอร์

Using Laser Plotter to Draw Graphic Picture

ผศ. พิพัฒน์ โชคสุวัฒนสกุล
เสมียน พรหมงาม ธาตรี นิกรรัมย์ จิระนาถ ขาวเมืองน้อย อวิรุทธิ์ โพธิชัย
ภาควิชาฟิสิกส์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

ABSTRACT-- There are many applications using laser as a tool for cutting, drilling or marking many kinds of non-metallic materials such as cloth, leather, acryric, wood and etc. We are now developing our software and hardware to control plotter as a laser marking tool. This laser plotter can not only use as an ordinary plotter but also can use to cut, draw or mark any pictures appear on the monitor screen. The quality of the marking picture is the same as the picture showing on the screen. Most of the processes are controlled by software instead of hardware. This laser plotter can mark with difference depth of any parts of a picture file by controlling plotter speed and/or laser power. The power of laser can be controlled by software with 256 levels. Of cause, our software and hardware must be used together to do these jobs.

KEY WORDS -- Laser, Plotter

ระบบวัดโฟโตรีเฟลกแตนท์สเปกโทรสโคปี้

Photoreflectance spectroscopy measurement system

ดร.จิติ หนูแก้ว, นางสาวทุติยภรณ์ ทิวาวงศ์, นายอภิชาติ สังข์ทอง, รศ. สุวรรณ คูสำราญ
ภาควิชาฟิสิกส์ประยุกต์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง และ
หน่วยปฏิบัติการวิจัยและพัฒนาเทคโนโลยีอิเล็กโทรออปติกส์
ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ

ABSTRACT-- The objective of this research is to construct the prototype of room-temperature photoreflectance (PR) spectroscopy measurement system for studying the energy band structures of semiconductor and semiconductor heterostructures. The system is set up on an optical table and modulation light is provided by a 3 mW He-Ne laser. The chopped laser light is irradiated onto the thin film sample. Light from a 100 W tungsten lamp passed through a 7 cm monochromator, acts as a probe light. The reflected probe light from the sample is detected for each wavelength from 500 nm to 2200 nm. The detected signal has two parts: The ac part measured by the lock-in amplifier synchronize to the modulating frequency is related to the change in reflectivity, dR, while its dc part is related to the reflectivity, R. Using a computer for control system and data acquisition, a spectrum of dR/R versus photon energy can be obtained.

KEYWORDS: photoreflectance, spectroscopy, energy band structures, semiconductor

พัฒนาโปรแกรมสำเร็จรูปสร้างสารานุกรม

Encyclopedia Building Software Development

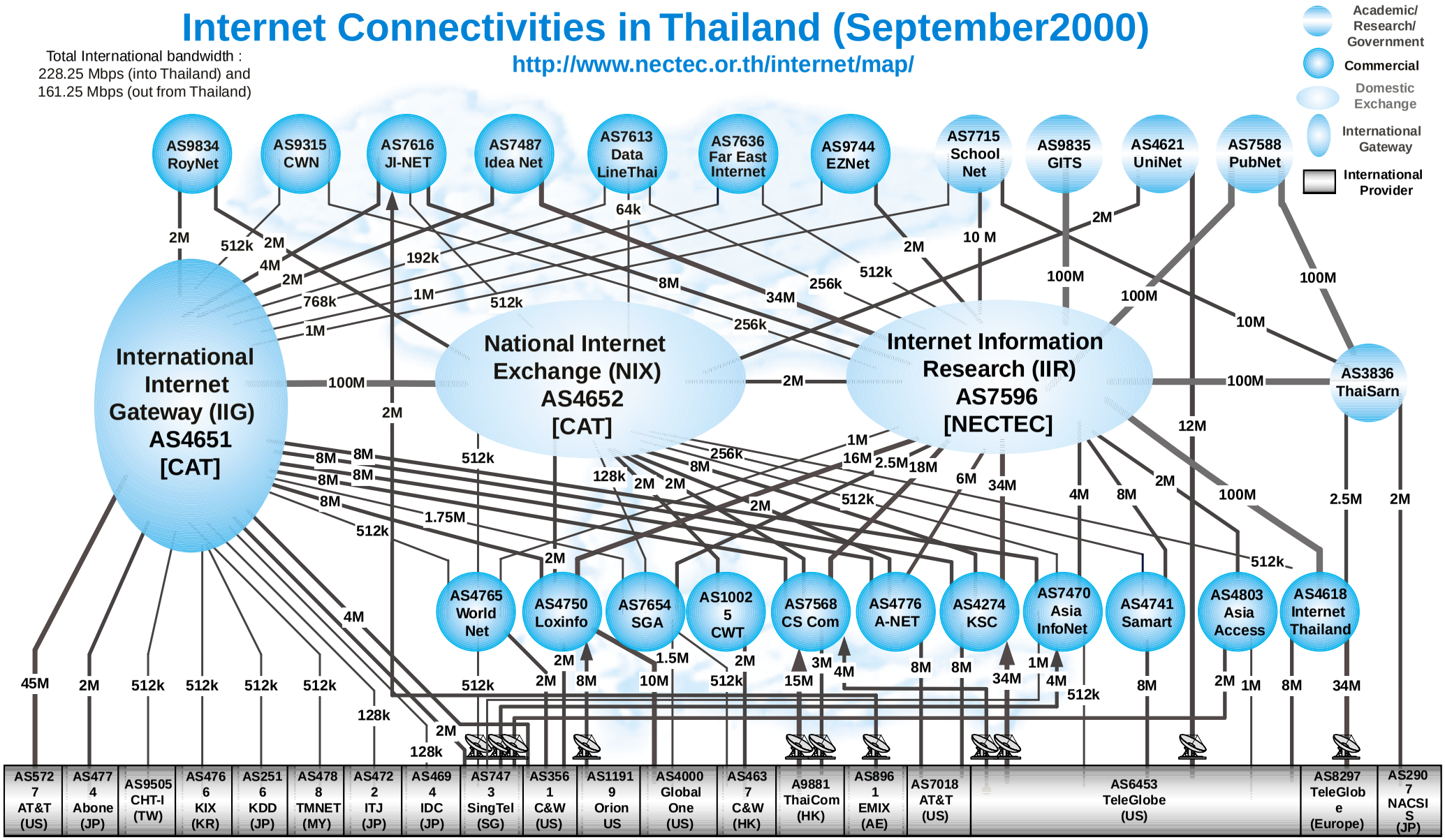
ผศ. กลชาญ อนันตสมบูรณ์ นางสาวกาลัญญู ปิยะสันติ์ นางสาวศรีนวล ฟองมณี
อาจารย์ สถาบันราชภัฏเชียงราย

ABSTRACT-- Nowadays, there are a small number of Computer-Aided Instruction (CAI) software in Thailand. Besides, most software was developed for specific purposes. The Encyclopedia Building software is a CAI, which can be created by the users themselves. Users can define key words, contents, and images used to illustrate the contents so that the students can use them as a referenced encyclopedia. User, moreover, can also define secondary key words appeared in any contents of the primary key words.

Internet Connectivities in Thailand (September2000)

http://www.nectec.or.th/internet/map/

Total International bandwidth :
228.25 Mbps (into Thailand) and
161.25 Mbps (out from Thailand)



DISCLAIMER

Chart Date: 2000-09-01

InetGTS ThaiSam