

ศักยภาพของเทคโนโลยี การแปลงเสียงเป็นข้อความ เพื่อจัดทำคำบรรยายแทนเสียง

ชัย วุฒิวิวัฒน์ชัย





Dragon Dictation
ฟรีแอปโดย Nuance ที่รองรับ
เสียงพูดภาษาไทย (2555)



Jibbigo Voice Translator
แอปโดย Facebook ที่รองรับ
การแปลเสียงพูดภาษาไทย
(2555)



Google Voice Search
ตั้งเป้าเปิดตัวบริการที่รองรับ
เสียงพูดภาษาไทย (ปลาย 2556)

iPhone 4s



2555



2558



Google Voice Search
รองรับเสียงพูดภาษาไทย
(2556)

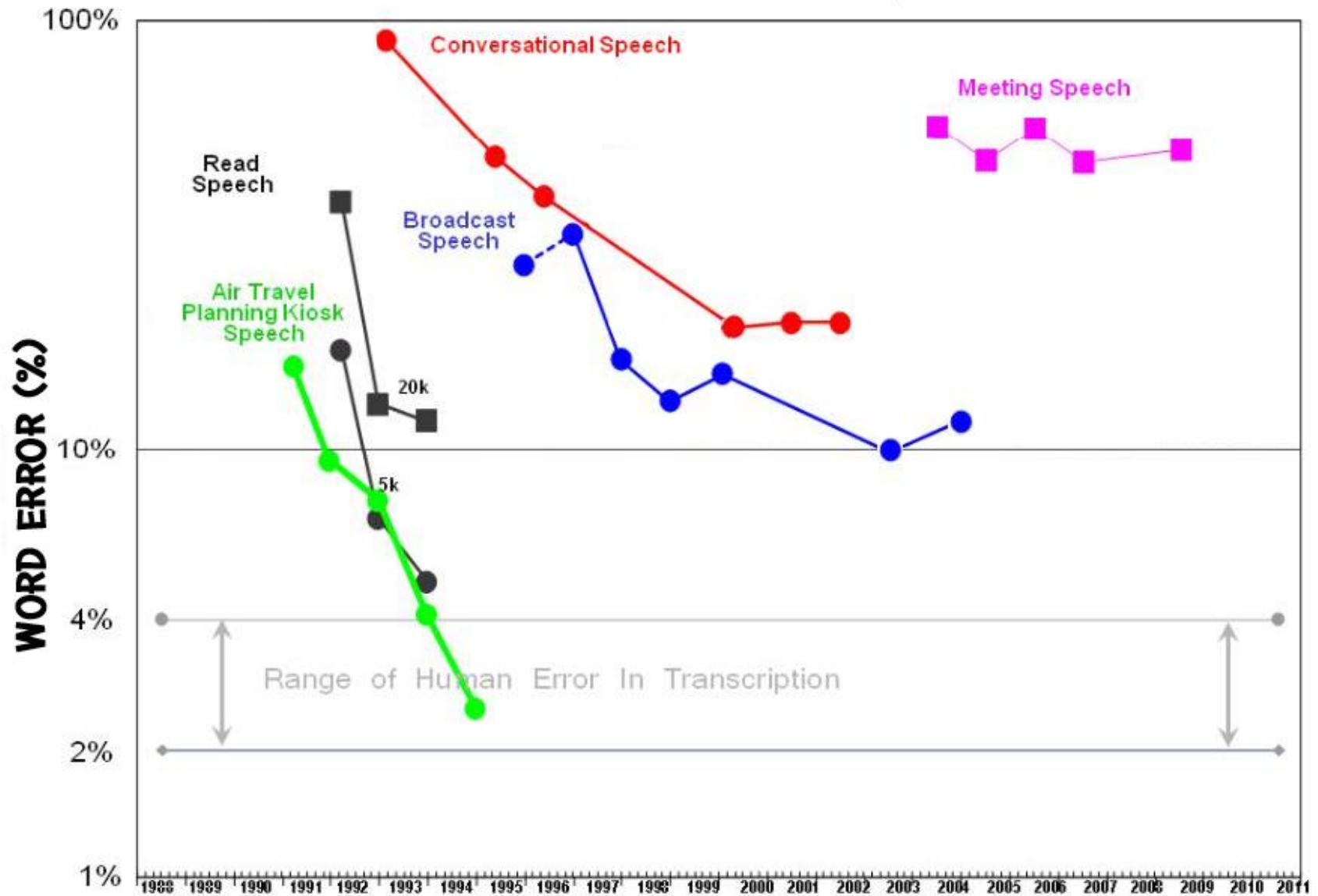


Apple iOS 8
VoiceOver พร้อม
Thai dictation (2557)



Skype Translate Beta
Speech translation
รองรับภาษาไทย (2557)

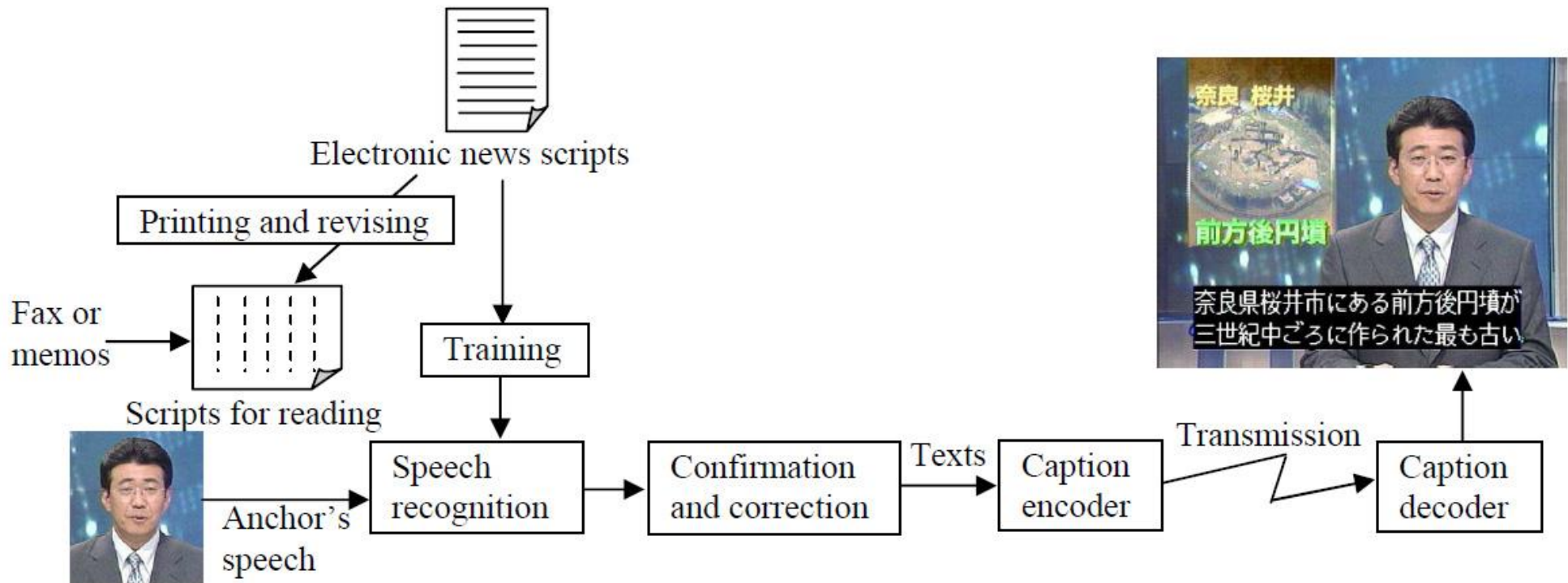
ประสิทธิภาพของระบบรู้จำเสียงพูด



Modified from NIST SST Benchmark, 2009

NHK Subtitling System

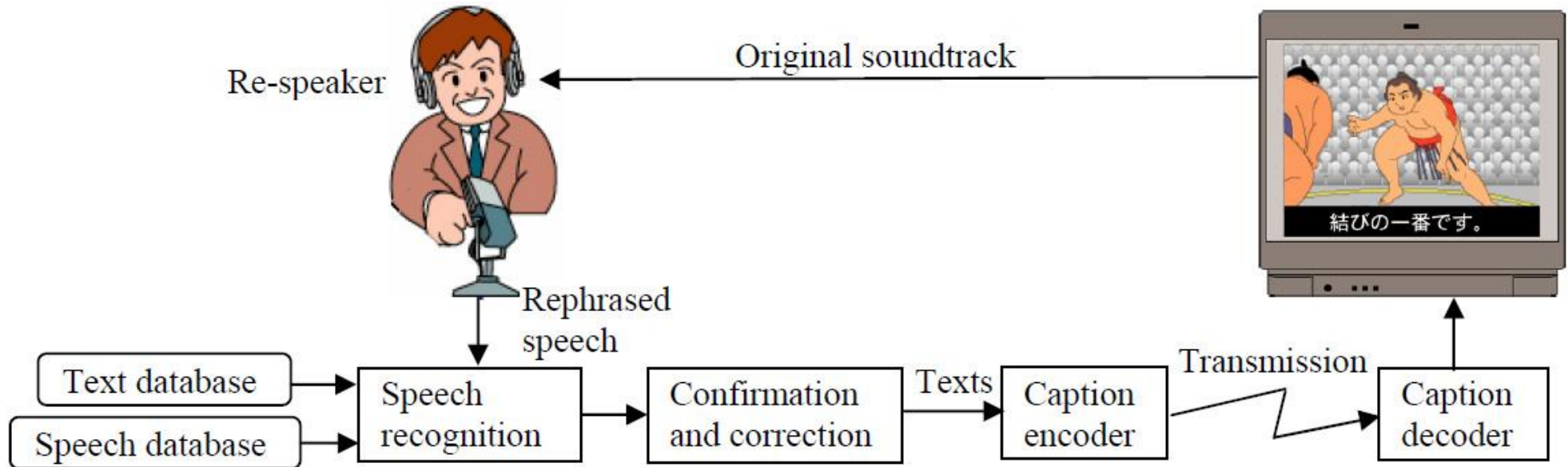
Imai et al. 2004. *Speech Recognition for Subtitling Japanese Live Broadcasts*. ICA 2004, Vol.I, pp.165-168.



- ความถูกต้องของการแปลงเสียงผู้ประกาศข่าวที่อ่านข่าวตาม script ได้ 98%
- ระบบมี delay ไม่เกิน 2 วินาที และใช้เวลาอีก 5 วินาทีในการตรวจสอบและแก้ไข รวมระยะเวลา delay ก่อนออกอากาศไม่เกิน 10 วินาที
- ความถูกต้องหลังจากแก้ไข ไม่ต่ำกว่า 99%

NHK Subtitling System: Re-speaking

Imai et al. 2004. *Speech Recognition for Subtitling Japanese Live Broadcasts. ICA 2004, Vol.I, pp.165-168.*



- ใช้ในกรณี Live program ประเภทบันเทิงและกีฬา ซึ่งมักใช้เสียงพูดแบบเป็นธรรมชาติ มีอารมณ์ และมีบรรยากาศหลากหลาย
- กรณีพูดซ้ำตามเสียงที่ได้ยิน (Repeating) ความถูกต้องประมาณ 93%
- กรณีพูดแบบเรียบเรียงใหม่ (Rephrasing) ความถูกต้องประมาณ 95%
- รวมระยะเวลา delay ก่อนออกอากาศไม่เกิน 8 วินาที
- ได้รับ positive feedback มากกว่าการถอดความตรง



โครงการระบบรับรู้จำเสียงพูดภาษาไทย 2556 – 2559

เทคโนโลยีจากต่างชาติ

- ตลาดเท่านั้นที่เป็นตัวจุดให้เกิดการพัฒนา
- มีข้อจำกัดอย่างมากในการปรับปรุงและให้บริการ





PARTY

สรุปคุณลักษณะปัจจุบัน



ไม่จำกัดเนื้อหา

ครอบคลุมคำศัพท์
มากกว่า 140,000 คำ



ถูกต้อง 80%

ภายใต้สภาวะแวดล้อม
Smartphone data channel



ตอบสนองเร็ว

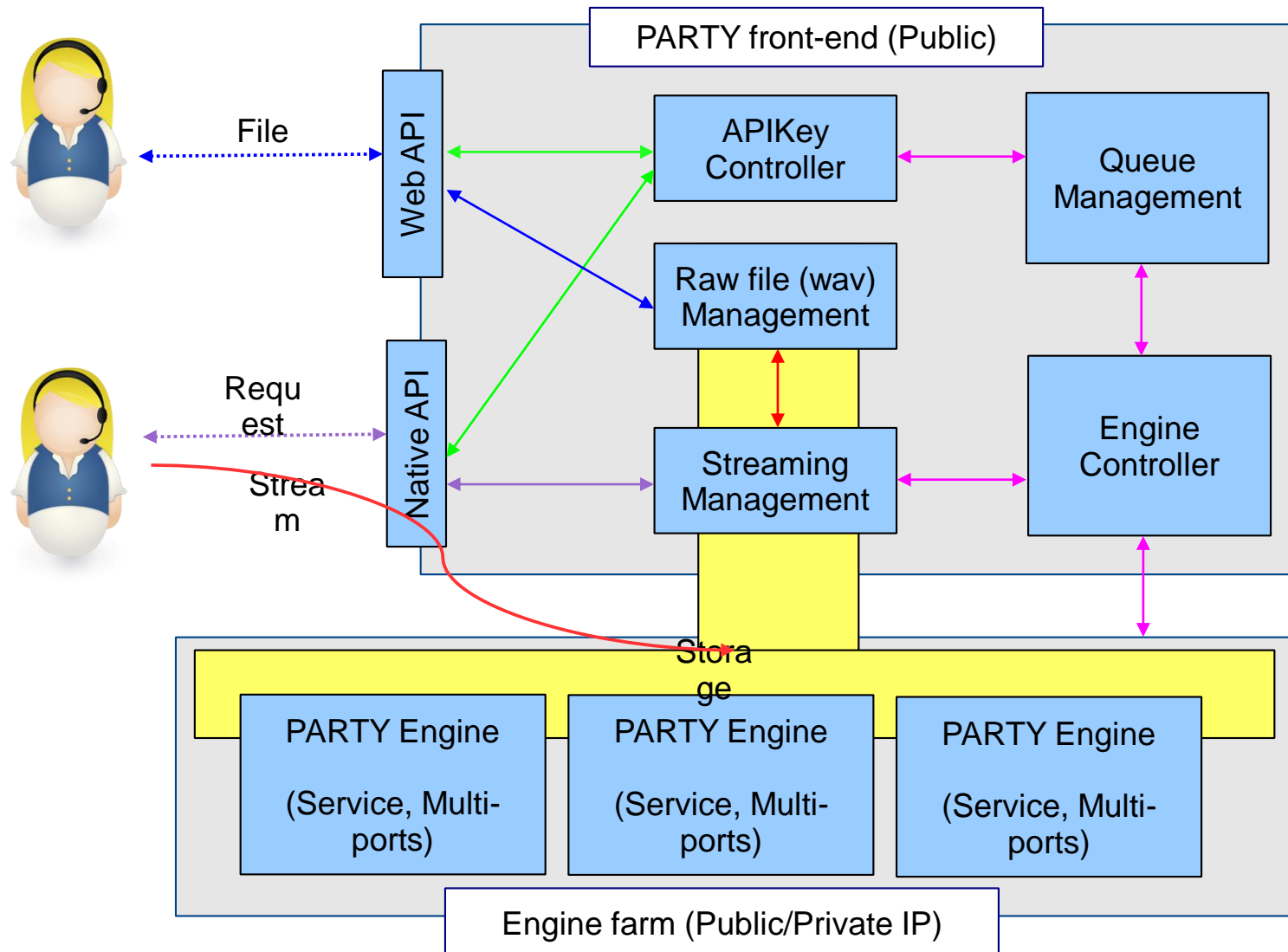
ให้คำตอบภายใน
1.5 เท่าของเวลา
เสียงอินพุต



ปรับแต่งได้

เว็บบริการแบบปรับแต่งได้
ตามความต้องการผู้ใช้

ประสิทธิภาพระบบ เว็บบริการขยายได้



Application to Audio Transcription



photo by www.fnfkorea.org



photo by www.thailandobservers.com



photo by www.utdallas.edu

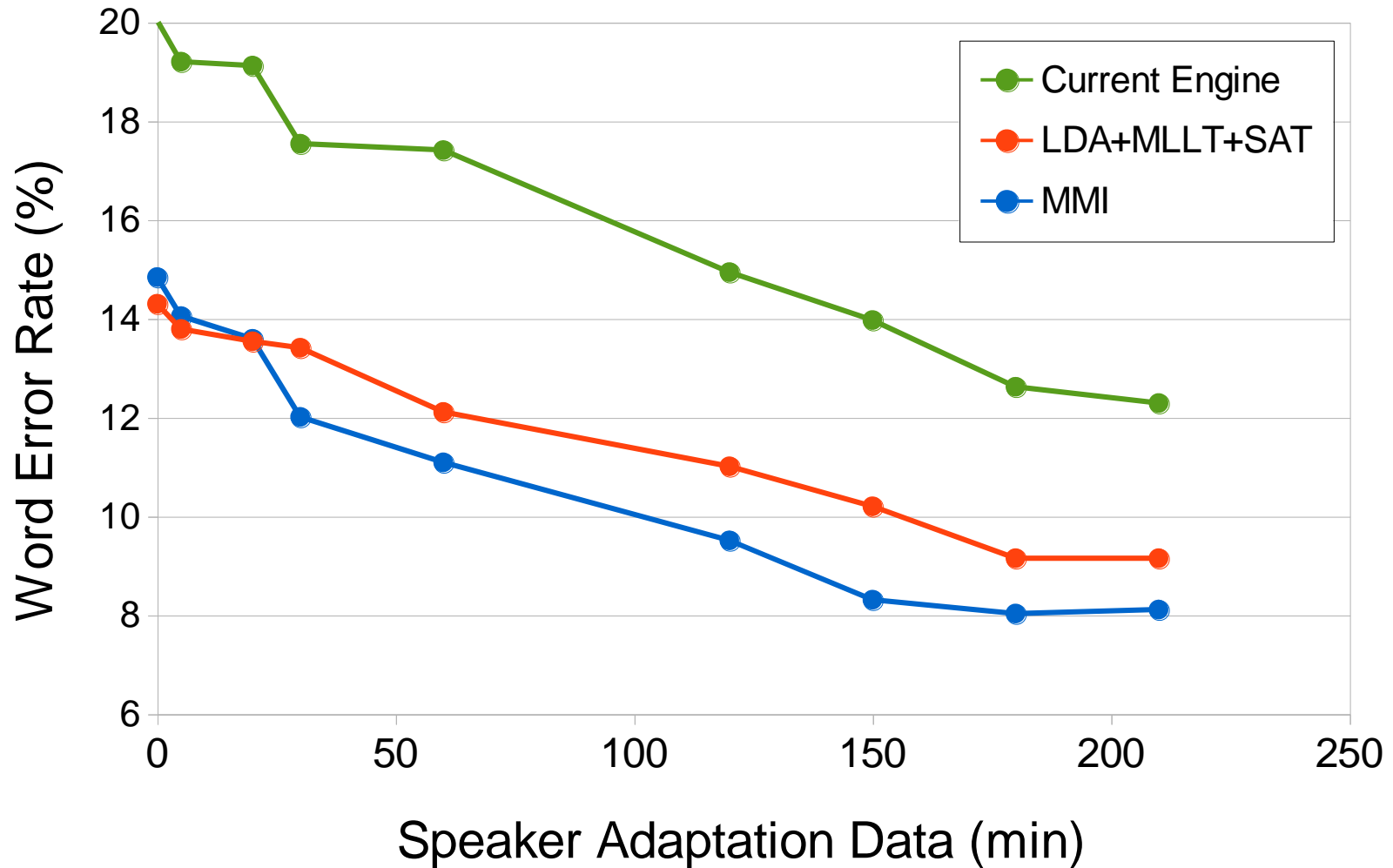


www.longtailvideo.com



www.Thai-Blogs.com

การปรับให้รองรับเสียงเฉพาะบุคคล Speaker Adaptation



การปรับให้รองรับ เสียงเฉพาะบุคคล

- ระบบตั้งต้นปรับด้วยเสียงพูด
จากที่ประชุมรัฐสภายาว 3.5 ชม.
- ทดสอบด้วยเสียงจากวิดีโอคลิป
ที่ได้จาก youtube



www.youtube.com/watch?v=Reejmlo8Jyc

พี่น้องประชาชนที่เคารพรักทุก **ท่าน**ครับ บ้านเมืองของเราวันนี้จะเดินต่อไปในสภาพแบบนี้ต่อไป **ไม่ได้**ครับ ทุกวันนี้พี่น้องประชาชนทุกคนเดือดร้อนจากสภาพเศรษฐกิจที่ชะงักงันของแพงพี่น้องเกษตรกรชาวนา **รอรอเงิน** จำนำข้าวพี่น้องจำนวนมาก **เมื่อไหร่**เมื่อไหร่กับการทุจริตกับการใช้อำนาจที่ไม่ถูกต้อง **ในทาง**การเมืองพี่น้องทุกคนเครียดกับความขัดแย้งทางการเมืองที่มีความรุนแรงมีกา **ก**ที่ จะ มอง **เห็น**ทางออก **ได้**ที่ สำคัญที่สุดทุกวัน **มี**นี้พี่น้องกังวล **ถูกต้อง 87%** เหตุการณ์ความรุนแรงจะมีมากขึ้น อาจเกิดถึงขั้นที่เราเรียกว่าเป็นสงครามกลางเมือง **เรา**หรือจะไปจบลงที่การปฏิวัติรัฐประหารซึ่งเป็นสิ่งที่ไม่มีใคร **รัก** **ท่าน** **ปรารถนา** วันนี้จึงไม่ใช่เวลาที่จะมาถกเถียงกันอีกต่อไป ไม่ใช่เวลาที่ **จังหวัด**จะมา โทษกันไปกัน **มาว่า** **เรา** เป็นความผิดของใคร **ของ** **เพราะ** ทุกคนล้วนมีส่วน **ให้** ประเทศ **มี** **มา** อยู่ตรงนี้รวมทั้งผมและพรรคประชาธิปัตย์ด้วย

Demo Application สำหรับงานถอดความ

File Text — Menu Bar

แถบเวลาและข้อความ

หน้าต่างแสดงเวลาเริ่มต้น
เวลาสิ้นสุด และข้อความ

หน้าแสดงภาพ Video

แถบ Player Control

StartTime	EndTime	Text
00:00:00.00	00:00:00.864	หยุด พัก กระหม
00:00:00.864	00:00:02.624	ที่ บ้าน ว่า ขนลุก กับ ต้อนรับ
00:00:02.624	00:00:06.688	บรรดา บุตร และ นักกีฬา ผู้ วิจารณ์ อภิปรายญา ทีมชาติไทย นะ ครับ
00:00:06.688	00:00:14.688	จำนวน ห้า สิบ สาม คน ก็ เดินทาง ไป แข่ง ชัน โดย เชิญ โอลิมปิก โลก
00:00:14.688	00:00:16.320	นาม ประเทศ อเมริกา
00:00:16.320	00:00:20.416	ไน วง ที่ เสนอ กรกฎาคม กุนซ้อ สี ห้า หก ร้อย ห้า สิบ แปด
00:00:20.416	00:00:24.608	ผม ถือ ว่า ไข่ แล้ว ก็ มนุษย์ นั้น ประจำ ก็ จะ มี คุณค่า ที่สุด ของ ชา
00:00:24.608	00:00:28.544	แล้ว การ กีฬา เป็น เครื่องมือ เพื่อ จะ ถูก ฝน ร่างกาย และ จิตใจ ใ
00:00:28.544	00:00:29.728	ทุก คน เข้มแข็ง
00:00:29.728	00:00:31.232	ได้ แล้ว ทุก คน นั้น ถือ ว่า เป็น
00:00:31.232	00:00:33.440	ตัวแทน ประเทศ เป็น ทูต พิเศษ นะ ครับ
00:00:33.440	00:00:36.416	ใน การ ที่ ฝรั่งเศส ก็ โอเค บาง ครั้ง กีฬา
00:00:36.416	00:00:39.424	รัฐบาล ว่า ลูกเรือ คน ทุก กลุ่ม อย่าง เท่าเทียม กัน
00:00:39.424	00:00:42.464	ผม ก็ คอย ดู ก็ ได้ ทำ หน้า ที่ ของ ดน อย่าง เต็มที่
00:00:42.464	00:00:43.456	ประสบ ความสำเร็จ
00:00:43.456	00:00:44.608	สม ชื่อ นุง
00:00:44.608	00:00:45.760	หวัง ไว้ ทุก คน นะ ครับ
00:00:45.760	00:00:52.00	ลูก กอล์ฟ ริเริ่ม หรือ การ ต่อต้าน อังกฤษ หรือ เป็น วาระ แห่งชาติ ที่ รั
00:00:52.00	00:00:55.776	ดู จาก การ ทูจริต คอร์ปชั่น เต็ม นะ ค่ะ คือ เรื่องริ่ง ของ ประเทศ
00:00:55.776	00:00:58.176	ที่ คอย ทำร้าย สังคม ไทย ของ เรา
00:00:58.176	00:01:03.104	เป็น ราคา ที่ สะท้อน ถึง วิฤกการณ์ ด้าน คุณธรรม จริยธรรม ของ คน ใ

Scroll Bar

the Special Olympics World Summer Games 2015 from July 21 to August 4.

ไทย 7/28/2015 4:53 PM

สรุป

- การเติบโตของเทคโนโลยีการแปลงเสียงพูดเป็นข้อความ โดยเฉพาะสำหรับภาษาไทยในระยะ 5 ปีที่ผ่านมา
- การทดสอบภาคสนามเพื่อเปรียบเทียบให้เห็นความคุ้มค่าของการใช้เทคโนโลยีกับการถอดความด้วยวิธีการอื่น
- ความเข้าใจธรรมชาติของเทคโนโลยีที่ต้องการการพัฒนาเติบโตไปพร้อมกับการใช้งานในช่วงเวลาที่เพียงพอ
- วิสัยทัศน์และการผลักดันของหน่วยงานที่เกี่ยวข้องต่อการใช้เทคโนโลยีเพื่อให้บริการสาธารณะที่ทั่วถึงและเท่าเทียม