บทที่ 2.

ประวัติการค้นคว้าเกี่ยวกับเรื่องเสียง

2.1 ประวัติการค้นคว้าเกี่ยวกับระบบตรวจรู้เสียง

การค้นคว้า พัฒนาเกี่ยวกับระบบตรวจรู้เสียงนี้ได้กระทำมานานกว่า30 ปีแล้ว และได้มีการสร้างระบบและทดลองต่างๆ มากมาย บางส่วนของการวิจัยเหล่านี้ได้แก่[4]

ในปี ค.ศ.1952 Davis,Wiren และ Biddulph จาก Bell laboratory ได้สร้างระบบตรวจรู้ ซึ่งสามารถตรวจรู้เสียงพูดตัวเลขได้ 10 ตัว สามารถจำคำพูดจากผู้พูด 1 คน ได้ถูกต้อง 100% โดยผู้พูดพูดผ่านโทรศัพท์ ความถูกต้องจะลดลงต่ำกว่า 50% ในกรณีที่เปลี่ยนคนพูด

ในปี ค.ศ.1956 Wiren และ Stubbs จาก Northeastern University ได้สร้างเครื่องตรวจรู้แยกประเภทของการออกเสียง เช่น คำเสียงก้อง(voiced) คำเสียงไม่ก้อง(unvoiced) คำที่มีเสียงกัก(stop) และเสียงเสียดแทรก(fricative) เสียงแหลม(acute) เสียงต่ำ(grave) เป็นต้น ระบบสามารถตรวจรู้ได้อย่างถูกต้อง 94% สำหรับผู้พูด 24 คน โดยจำเสียงสระ ในปีเดียวกัน Olson และ Belar จาก R C A ได้สร้างระบบสามารถตรวจรู้คำ 1 พยางค์ได้ 10 คำ มีความถูกต้อง 98% สำหรับผู้พูด 1 คน โดยผู้พูดจะต้องระวังการออกเสียงแต่ละพยางค์ในการพูดประโยค 1 ประโยค และจะต้องหยุดระหว่างคำ ระบบใช้วงจรกรองแถบความถี่ 8 วงจร และแบ่งการวิเคราะห์ออกเป็น 5 ช่วงเวลาต่อคำ ซึ่งจะได้แมทริก(Matrix) ขนาด 8x5 ใช้ในการวิเคราะห์ แต่ละสมาชิกของแมทริกจะมีค่า 1 หรือ 0 ขึ้นกับค่าพลังงานแต่ละแถบความถี่ จะมีค่ามากกว่าหรือน้อยกว่าระดับที่กำหนดไว้

ในปี ค.ศ.1960 Danes และ Mathew ได้สร้างระบบตรวจรู้เลข 10 จำนวน โดยใช้เครื่องวิเคราะห์สเปคตรัม 17 ช่องความถี่ จับสัญญาณและบันทึกข้อมูลลงเทปแม่เหล็ก จากนั้นนำข้อมูลเข้าสู่คอมพิวเตอร์เพื่อทำการวิเคราะห์

ด้านความถี่และเวลา แล้วเก็บบันทึกไว้เป็นแบบอ้างอิง คำพูดใหม่ที่เข้ามาจะถูกเปรียบเทียบโดย ขบวนการครอสโคริเลชั่น(crosscorrelation process) กับแต่ละรูปแบบที่เก็บไว้ ซึ่งระบบจะเลือกรูปแบบที่คล้ายกันมากที่สุดกับรูปแบบของคำพูดใหม่ การเปรียบเทียบกระทำโดยใช้การนอมัลไลซ์(normallization) และโดยไม่นอมัลไลซ์

ในปี ค.ศ.1966 King และ Tunis ได้สร้างเครื่องตรวจรู้คำ (word recognition) ใช้เครื่องวิเคราะห์สเปคตรัม 15 ช่อง และคอมพิวเตอร์ช่วยในการวิเคราะห์ โปรแกรมคอมพิวเตอร์ทำหน้าที่ตรวจจับยอดที่เกิดขึ้นในแต่ละแถบความถี่ โดยหาตำแหน่งที่เกิด สร้างเป็นแมทริกซ์ความถี่-เวลา (frequency-time matrix)ซึ่งสมาชิกจะมีค่าเป็น 1 ถ้าในแถบความถี่ ณ เวลานั้นเกิดยอดคลื่นขึ้น และเป็น 0 ถ้าไม่มียอด ผลลัพท์ให้ความผิดพลาด 0.2% เมื่อตัวอย่างเสียงเก็บจากผู้พูดคนเดียวกัน และผิดพลาดมากที่สุด 2.5% การสอนให้ระบบรู้จัก จะใช้คำพูด 15 คำ แต่ละคำจะให้ตัวอย่าง 35 ถึง 50 ครั้ง การทดลองให้ผู้ทดลองพูดคำต่างๆ 90 ถึง 115 ครั้ง ผลลัพท์จะผิดพลาดมากที่สุดเมื่อตัวอย่างเสียงพูดเก็บจากคนหนึ่งและใช้ผู้พูดอีกคนหนึ่ง ความผิดพลาดจะเกิดขึ้นประมาณ 46% แต่เมื่อการเก็บตัวอย่างเสียงกระทำจากคนทั้ง 2 คน ความผิดพลาดจะมีเพียง 0.8%

ในปี ค.ศ.1966 Fraipont ได้สร้างเครื่องตรวจรู้คำ โดยใช้วงจรกรองแถบความถี่ 10 แถบ กรองความถี่ในช่วง 300 Hz. ถึง 3000 Hz. โดยจับเฉพาะแนวยอดคลื่น(envelope) และใช้วงจรกรองแถบความถี่อีก 2 วงจรสำหรับความถี่สูงกว่า 3000 Hz. การสุ่มตัวอย่างสัญญาณใช้เวลา 500 msec. การสุ่มคำพูดกระทำ 192 ครั้ง ข้อมูลถูกบันทึกลงบนเทปกระดาษ แล้วป้อนเข้าสู่คอมพิวเตอร์ ซึ่งทำการวิเคราะห์โดยใช้ ลิเนียร์ ดิซิชั่น ไฮเปอร์เพลน(linear decision hyperplane) ซึ่งจะถูกปรับให้เหมาะสม ระหว่างการสอนระบบ ระบบถูกสร้างให้ตรวจรู้ตัวเลขได้ 10 ตัว และคำอื่นๆอีก 5 คำ ข้อมูลตัวอย่างเก็บจากคน 24 คน แต่ละคนจะพูดคำหนึ่งคำเพียงครั้งเดียว หลังจากสอนระบบเรียบร้อยแล้ว ระบบจะสามารถตรวจรู้ได้ถูกต้อง 99% สำหรับผู้พูดกลุ่มเดิม และถูกต้อง 87% เมื่อเปลี่ยนกลุ่มผู้พูด

ในปีพ.ศ.2526 สุพงศ์ เภกะนันทน์ ได้วิจัยการตรวจรู้เสียงพูดตัว เลข ในภาษาไทย โดยใช้เทคนิคการวิเคราะห์สัญญาณตามความถี่ และได้ผลในการตรวจรู้ถูกต้อง 60 % แต่ถ้าผู้พูดได้รับการฝึกฝนเป็นพิเศษจะสามารถตรวจรู้ได้ถูกต้องมากกว่า 90 %

ในปี ค.ศ.1985 วีระ ริ้วพิทักษ์ ได้วิจัยการตรวจรู้พยางค์ในภาษาไทยโดยใช้สัมประสิทธิ์การสะท้อนกลับ เป็นพารามิเตอร์ที่ใช้ในการเปรียบเทียบ ได้ผลถูกต้อง 89 % และได้ใช้พารามิเตอร์ความถี่มูลฐานในการแยกวรรณยุกต์ ซึ่งได้ผลถูกต้องในการตรวจรู้ถึง 92.8 %

2.2 ประวัติการค้นคว้าเกี่ยวกับระบบสัทสัง เคราะห์

การศึกษาค้นคว้าทางด้านสัทสัง เคราะห์เสียงพูดของคนเรา เริ่มมาตั้งแต่สมัยโบราณ แต่ที่เป็นการศึกษาอย่างเป็นระบบนั้น เริ่มมาเมื่อประมาณ 200 ปีที่แล้ว โดยเริ่มพัฒนามาเป็นลำดับดังนี้[5]

ในปี ค.ศ.1770 Kratzenstein ได้ผลิตเครื่องสัง เคราะห์เสียงขึ้นมา เป็นลักษณะของการนำตัวก่อให้เกิดกำทอน(resonators) หลายๆตัวมาต่อกัน

และในปี ค.ศ. เดียวกันนั้นเอง Wolfgang Von Kempelen ได้สร้าง เครื่องสัง เคราะห์เสียงขึ้นมาซึ่งมีลักษณะที่ซับซ้อนมากกว่าของKratzenstein และต่อมาระบบนี้ได้สร้างขึ้นอีกครั้งโดย Wheatstone ระบบดังกล่าวเป็นการสร้างส่วนประกอบต่าง ๆ ให้มีหน้าที่เลียนแบบอวัยวะที่ใช้ในการออกเสียงของมนุษย์ มีรูปร่างดังรูปที่ 2

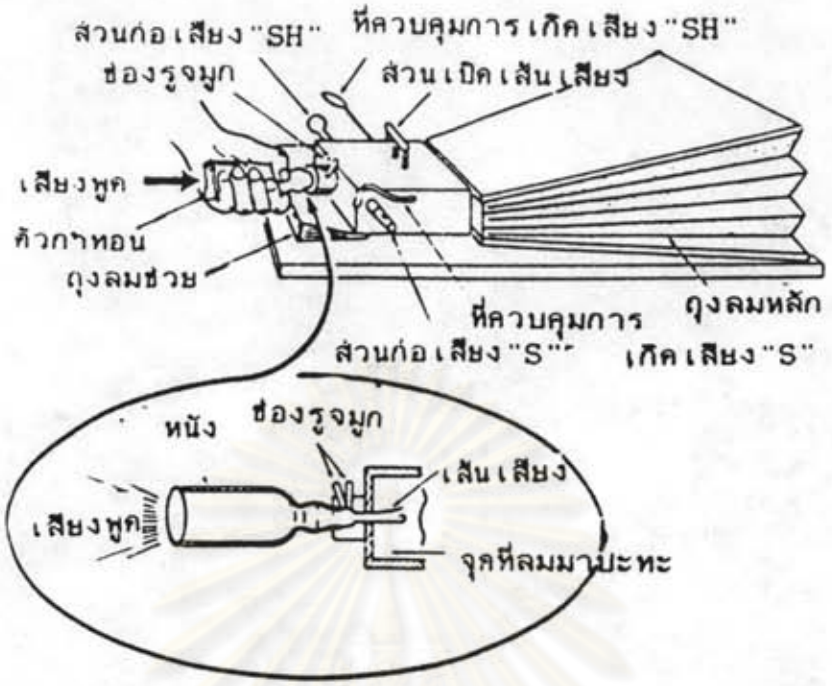
ในปี ค.ศ.1876 Alexander Graham Bell ได้พัฒนาระบบของ Wheatstone ขึ้นโดยใช้วัสดุจำพวก ไม้, ยาง และผ้าม โดยพยายามจำลองรูปแบบของอวัยวะที่ใช้ในการออกเสียง เช่น ริมฝีปาก, แก้ม, ลิ้น, เส้นเสียง เป็นต้น

ในปี ค.ศ.1922 J.Q.Stewart ได้สร้าง เครื่องสัง เคราะห์เสียงขึ้นจากระบบอิเลคโทรนิคส์ โดยระบบนี้จะมีตัวที่ทำให้เกิดเสียงอยู่ตัวหนึ่ง เพื่อใช้แทนการทำงานของ เส้นเสียง และจะมีตัวความต้านทาน, ตัว เก็บประจุ และตัว เหนี่ยวนำที่จะก่อให้เกิดการกำทอน ซึ่งระบบดังกล่าวมีลักษณะดังรูปที่ 3

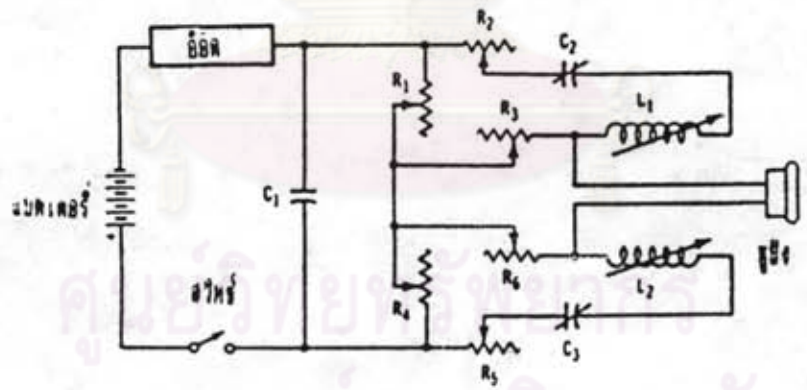
ในปี ค.ศ. 1937 RR. Riesz ได้สร้างเครื่องสังเคราะห์เสียง โดยอาศัยหลักการคล้ายการทำงานของหรัมเปต โดยมีลิ้นที่จะต้องควบคุมทั้งหมด 10 ลิ้น ดังรูปที่ 4 และจากรายงานบอกว่า เครื่องดังกล่าวสามารถที่จะสังเคราะห์เสียง "cigarette" ได้

ในปี ค.ศ. 1939 RR. Riesz, H. W. Dudley และ S. A. Watkins ได้ร่วมกันสร้างเครื่องสังเคราะห์เสียง โดยอาศัยหลักการทางอิเลคโทรนิคส์ และทางกลร่วมกัน ซึ่งระบบดังกล่าว จะมีส่วนที่ใช้สังเคราะห์เสียงก้อง(Voice) และเสียงไม่ก้อง(Unvoice) ขึ้น ซึ่งควบคุมโดย การบังคับคาน(Wrist Bar) หลังจากนั้นสัญญาณจะผ่านไปยังส่วนการกำหอน เพื่อกรองความถี่ 10 ช่วงความถี่ ซึ่งควบคุมโดยกลุ่มคานบังคับ(Key Board) โดยระบบดังกล่าวเรียกว่า "Voder" และมีลักษณะดังรูปที่ 5

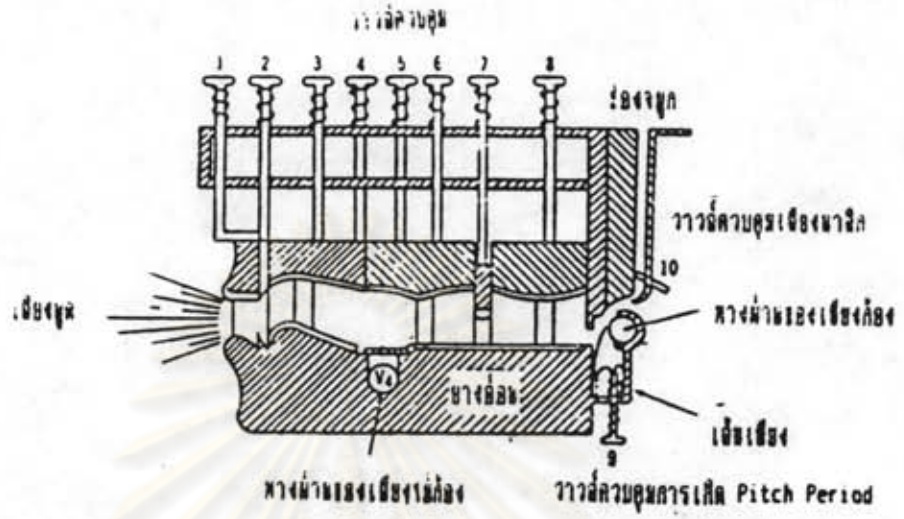
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



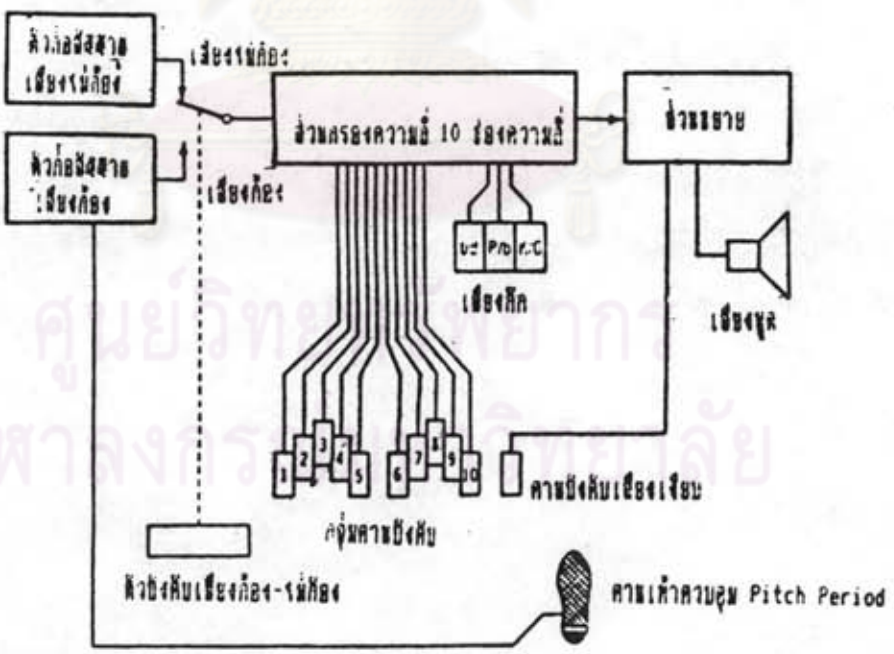
รูปที่ 2 ระบบการสร้างเสียงของ Wheatstone



รูปที่ 3 ระบบการสร้างเสียงของ J.Q.Stewart



รูปที่ 4 ระบบการสร้างเสียงของ RR. Riesz



รูปที่ 5 ระบบการสร้างเสียงของ RR. Riesz, H. W. Dudley, S. A. Watkins