บทที่ 4
ขั้นตอน ผล และวิเคราะห์ผลของการทดลอง

ในบทนี้จะแบ่งเป็น 3 ตอนคือ

ขั้นตอนการทดลอง ผลการทดลอง และการวิเคราะห์ผลการทดลอง

1. ขั้นตอนการทดสอบ

1 สำหรับกรณีทดสอบแบบของตัวอักษรไทยบนวินโดว์ เวอร์ชั่น 3.1 ไทยเอดิชั่น โดยแบบอักษรที่นำมาใช้คือ AngsanaUPC , CordiaUPC จะทดสอบที่ขนาด 12 14 16 18 20 22 24 point และ แบบอักษร BrowalliaUPC , DilleniaUPC , EucrosiaUPC , FreeciaUPC , IrisUPC , JasmineUPC , KodchiangUPC , LilyUPC จะทดสอบที่ขนาด 14 point โดยพิมพ์บทความตัวอย่างจากต้นฉบับด้วยโปรแกรม Microsoft Word 6.0 โดยให้จำนวนอักษรต่อบรรทัดเท่ากับต้นแบบ แล้วพิมพ์โดยใช้เครื่องพิมพ์เลเซอร์ที่มีความละเอียด 300 จุดต่อนิ้ว สำหรับกรณีทดสอบกับเอกสารจริงทำการสแกนจากเอกสารต้นฉบับจริงโดยตรง (ตัวอย่างภาพที่สแกนได้ดูแสดงดังรูป ก.1 ก.2 และ ก.3 ในภาคผนวก)

** บทความต้นฉบับเป็นบทความที่พบเห็นทั่วไปคือ หนังสือพิมพ์ วารสาร นิยาย ดังนั้นจำนวนอักษรต่างๆ ที่ทดสอบจึงเป็นตามที่เกิดขึ้นจริง*

2 สแกนด้วยสแกนเนอร์ความละเอียด 300 จุดต่อนิ้ว และเป็นขาวดำ (binary image / monochrome image) โดยใช้โปรแกรม Photo Styler 2.0 แล้วเก็บในรูปแบบ PCX

หมายเหตุ สำหรับการทดลองปกติจะพยายามสแกนให้เอียงไม่เกิน 1 องศาซึ่งไม่ต้องแก้ความเอียง แต่เมื่อต้องการทดสอบโปรแกรมส่วนแก้ความเอียงจะเอียงเอกสารก่อนสแกนแล้ววัดว่าเอียงเท่าไหร่เพื่อบันทึกผลการทดลอง

3 ให้โปรแกรมรู้จำทำการรู้จำแฟ้มภาพที่สแกน *โดยเลือกแบบสุ่มดังนั้นจำนวนอักษรทดสอบในแต่ละแบบและขนาดอักษรจะไม่เท่ากัน*

4 ตรวจสอบความถูกต้องใช้เกณฑ์นับดังนี้

- ตัดอักษรเดี่ยวผิดเป็น 2 อักษรนับผิด 1 อักษร
- ไม่ตัดอักษรที่ติด นับเท่าจำนวนอักษรที่ติด
- รู้จำตัวเดี่ยวผิดและนับจำนวนที่อักษรที่รู้จำผิด

นับเวลาการรู้จำโดยนับตั้งแต่เริ่มตัดคอลัมน์และบรรทัด

บทความที่ใช้

1. บทความต้นฉบับที่ใช้ทดสอบแบบอักษรของวินโดว์คือ
   - บทความจากหนังสือ "หลายชีวิต" หน้า 347 - 357 ของ ม.ร.ว. คึกฤทธิ์ ปราโมช พิมพ์เป็นพ็อกเก็ตบุ๊ค ครั้งแรก เมษายน 2525 (แสดงตัวอย่างภาพถ่ายเอกสารในรูปที่ ก.4)
   - ข่าวจากหนังสือพิมพ์ไทยรัฐ ฉบับวันเสาร์ที่ 12 ตุลาคม พ.ศ. 2539 หน้าที่ 23 เรื่อง "อย.เตรียมคุมเข้มสถานชะลอความแก่" (แสดงตัวอย่างภาพถ่ายเอกสารในรูปที่ ก.5)

2. บทความจริงที่ใช้สแกนโดยตรงคือ
   - ข่าวจากหนังสือพิมพ์ไทยรัฐ ฉบับวันอาทิตย์ที่ 22 ธันวาคม พ.ศ. 2539 หน้าที่ 23 เรื่อง " หวยล๊อก '051' " (แสดงตัวอย่างภาพถ่ายเอกสารในรูปที่ ก.6)
   - บทความจากหนังสือ "Advanced Thailand Geographic" ปีที่ 2 ฉบับที่ 17 พฤศจิกายน พ.ศ. 2539 หน้า 151-152 (แสดงตัวอย่างภาพถ่ายเอกสารในรูปที่ ก.7)
   - บทความจากหนังสือ "ผู้หญิงวันนี้" ฉบับที่ 62 ประจำเดือนกรกฎาคม พ.ศ. 2539 หน้า 216 - 218 เรื่อง "ผ่าตัดไขมันหน้าท้อง" (แสดงตัวอย่างภาพถ่ายเอกสารในรูปที่ ก.8)

2. ผลการทดสอบ

** หมายเหตุ ตารางที่ 4.1 - 4.3 4.5 และ 4.6 อักษรทดสอบเลือกแบบสุ่มจากบทความที่พิมพ์จากต้นฉบับ*

ตารางที่ 4.1 แสดงผลการทดสอบของอักษรที่พิมพ์ด้วยเครื่องพิมพ์เลเซอร์

แบบอักษร	ขนาดอักษร (point)	จำนวนตัวอักษรทดสอบ (ตัว)	จำนวนรู้จำผิด (ตัว)	ความถูกต้อง (%)	เวลาที่ใช้ (ตัวต่อวินาที)
AngsanaUPC	12	6714	293	95.6	67.1
	14	1356	23	98.3	56.5
	16	3351	74	97.8	41.3
	18	3029	65	97.8	35.6
	20	2203	47	97.7	22.3
	22	3575	44	98.7	21.6
	24	3464	70	97.9	19.6
CordiaUPC	12	3156	174	94.5	65.8
	14	4145	58	98.6	51.2
	16	3484	45	98.7	41.9
	18	2779	38	98.6	31.9
	20	4017	102	97.5	24.2
	22	3657	77	97.9	20.2
	24	3321	58	98.3	17.6
สรุป	-	48251	1168	97.6	36.4
*ถ้าไม่รวมขนาด 12 point	-	38381	701	98.2	31.5

ตารางที่ 4.2 แสดงความผิดพลาดของอักษรที่พิมพ์ด้วยเครื่องพิมพ์เลเซอร์ลักษณะต่างๆ

แบบอักษร	ขนาดอักษร	ความผิดพลาดของอักษรลักษณะต่างๆ เทียบกับความผิดพลาดทั้งหมด(%)				
	(point)	1 (บน)	2 (บน + กลาง)	3 (กลาง)	4 (กลาง + ล่าง)	5 (ล่าง)
AngsanaUPC	12	53.5	1.7	43.3	0.3	1.2
	14	69.5	0.6	25.7	0.2	4.0
	16	58.1	1.9	36.4	1.1	3.5
	18	54.6	0.8	39.2	0.4	5.0
	20	55.0	0.6	44.2	0.1	0.1
	22	49.2	0.4	50.0	0.1	0.3
	24	50.4	0.7	50.5	0.1	0.3
CordiaUPC	12	31.2	0.8	66.0	0.5	1.5
	14	50.0	0.8	48.5	0.2	0.5
	16	57.0	0.4	42.1	0.3	0.2
	18	56.1	0.4	43.1	0.3	0.1
	20	39.4	0.2	61.2	0.1	0.1
	22	39.2	0.4	61.2	0.1	0.1
	24	35.5	0.6	63.5	0.1	0.3

ตารางที่ 4.3 แสดงจำนวนอักษรที่ทำการทดสอบและที่ติดสำหรับอักษรแบบและขนาดต่าง ๆ

	AngsanaUPC				CordiaUPC			
	12	14	16-18	20-24	12	14	16-18	20-24
จำนวนอักษรทดสอบ(ตัว)	6714	1356	3380	9242	3156	4145	6263	10995
จำนวนอักษรที่ติด(ตัว)	510	65	44	0	258	174	88	22
% ที่ติด	7.6	4.8	1.3	0.0	8.2	4.2	1.4	0.2

ตารางที่ 4.4 แสดงผลการทดสอบจากเอกสารจริง

ชื่อหนังสือ	ใกล้เคียงกับอักษร	จำนวนตัวอักษรทดสอบ(ตัว)	จำนวนอักษรที่รู้จำผิด(ตัว)	% ความถูกต้อง	เวลาที่ใช้ (ตัวต่อวินาที)
ไทยรัฐ	AngsanaUPC 11 point	2375	397	83.3	78.4
ผู้หญิงวันนี้	CordiaUPC 14 point	2329	46	98.0	51.8
Advance Thailand Geographic	CordiaUPC13 point	3351	121	96.4	56.2

ตารางที่ 4.5 แสดงผลการทดสอบของอักษรที่พิมพ์ด้วยเครื่องพิมพ์เลเซอร์

แบบอักษร	ขนาดอักษร (point)	จำนวนตัวอักษรทดสอบ (ตัว)	จำนวนรู้จำผิด (ตัว)	ความถูกต้อง (%)	เวลาที่ใช้ (ตัวต่อวินาที)
BrowalliaUPC	14	4676	174	96.3	59.3
DilleniaUPC	14	2176	110	94.9	64.9
EucrosiaUPC	14	2896	52	98.2	61.1
FreesiaUPC	14	1538	86	94.4	59.4
IrisUPC	14	1472	130	91.2	61.6
JasmineUPC	14	1321	488	63.0	65.5
KodchiangUPC	14	1080	460	57.4	73.1
LilyUPC	14	ไม่สามารถทดสอบได้เนื่องจากอักษรติดมาก			

ตารางที่ 4.6 แสดงผลการรู้จำเอกสารที่เอียงด้วยมุมต่างๆ (เอกสารพิมพ์ด้วยแบบอักษร AngsanaUPC 14 point)

มุมที่เอียง (องศา)	เมื่อแก้ความเอียงแล้วเหลือความเอียง (องศา)	จำนวนอักษรทดสอบ	จำนวนอักษรผิดพลาด	อัตราการรู้จำ (%)
0	0	2400	42	98.3
+2	0.2	2400	53	97.8
-2	0.5	2400	85	96.4
+5	0.2	2400	130	94.6
-5	0.3	2400	183	92.3
+7	0.5	2400	235	90.2
-7	0.6	2400	212	91.2

หมายเหตุ มุมเอียง + คือตามเข็มนาฬิกา - คือ ทวนเข็มนาฬิกา

3. การวิเคราะห์ผลการทดลอง

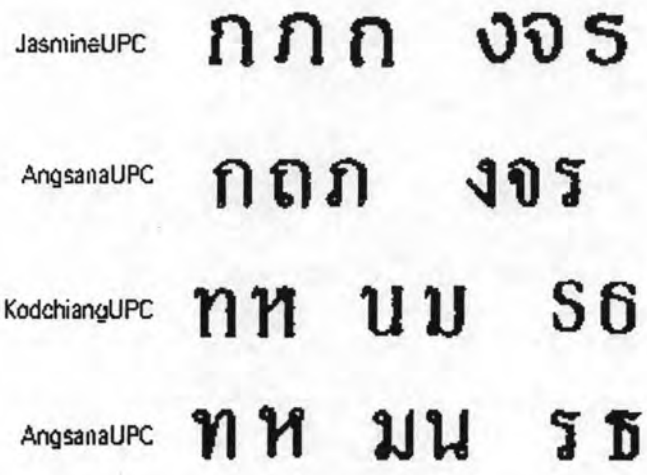
จากตารางที่ 4.1 จะเห็นได้ว่าอัตราความผิดพลาดที่ขนาดอักษร 12 point สูงกว่าขนาดอื่นมากเป็นผลมาจากตัดบรรทัดผิด อักษรติดและลักษณะบ่งความต่างบางอย่างหายไปเนื่องจากอักษรมีขนาดเล็ก เมื่ออักษรมีขนาดตั้งแต่ 14 point ขึ้นไปความผิดพลาดเริ่มคงที่ ได้ผลความถูกต้องประมาณ 97.4 % (ถ้าไม่รวมอักษรขนาด 12 point ตามขอบเขตการทดลองจะได้ความถูกต้องเป็น 98.2 %) เนื่องจากอักษรที่มีขนาดใหญ่กว่า 16 จะถูก Normalize มาเป็นขนาด 14 point จึงมีผลการรู้จำเหมือนอักษรขนาด 14 point ส่วนผลของอักษรติดมีผลน้อยเพราะอักษรขนาด 14 point ขึ้นไปมีอัตราการติดต่ำแสดงดังตารางที่ 3.3 ในบทที่แล้ว

สำหรับเวลาจะเพิ่มขึ้นตามขนาดของอักษร เนื่องจากถ้าหน้ากระดาษขนาดเท่ากันอักษรขนาดเล็กมีจำนวนอักษรมากกว่า แต่ใช้เวลาในการหาเส้นตัดคอลัมน์บรรทัดใกล้เคียงกันเพราะต้องนับจุดตามแนวดิ่งและแนวนอนเท่ากัน การดึงเกาะของอักษรก็เช่นกัน อักษรขนาดใหญ่ใช้เวลามากกว่าและเสียเวลาจากการ Normalize ด้วย

จากตารางที่ 4.2 จะเห็นได้ว่าน้ำหนักความผิดพลาดของ 2 แบบอักษรมีลักษณะต่างกันเนื่องจากการออกแบบการวางตัวอักษรและแบบตัวอักษรแต่ที่เหมือนกันคือความผิดพลาดที่ 12 point จะต่างจากขนาดอื่น อธิบายได้ว่าอักษรขนาด 12 point ความผิดพลาดส่วนใหญ่มาจากส่วนตัดบรรทัดตัดผิด อักษรติด และลักษณะบ่งความต่างบางอย่างที่หายไป เช่น หยักของอักษร “ต” “ซ” เป็นต้น ซึ่งแบบอักษร CordiaUPC ผิดที่ระดับกลางมากกว่าเนื่องจากอักษรระดับกลางติดกัน เช่น “เม” และอักษรระดับบนติดกันเองน้อยกว่าแบบอักษร AngsanaUPC ส่วนที่ลักษณะบ่งความต่างหายไปมีความผิดพลาดใกล้เคียงกัน เมื่ออักษรขนาดใหญ่ขึ้นผลความผิดพลาดจะไม่ได้มาจากอักษรติดอีกต่อไป จะเป็นจากส่วนรู้จำตัวเดี่ยวอย่างเดียว สำหรับอักษระดับกลาง +บน กลาง+ล่าง และล่างนั้น ที่มีน้ำหนักความผิดพลาดต่ำ เหตุผลหลักคือมีจำนวนสมาชิกน้อย และเป็นอักษรที่พบไม่บ่อยเช่น “ฎ” “ฤ” “ฬ” เป็นต้น

จากตารางที่ 4.3 แสดงอัตราอักษรที่ติดกัน เห็นได้ว่าอักษรขนาด 12 point มีการติดกันมากที่สุด สำหรับอักษร AngsanaUPC ไม่มีอักษรติดกันเลยที่ 20 point ขึ้นไป สำหรับอักษร CordiaUPC ที่ 20 point ขึ้นไปนั้นมีแบบอักษรติดที่พบแบบเดียวคืออักษรระดับ 1 ติดกับระดับ 2 เช่น “ ป” และเป็นต้นเหตุของความผิดพลาดเนื่องจากหางของ “ป” ไปทับกับไม้หันอากาศ

เมื่อดูผลจากตารางที่ 4.1 ร่วมกับตารางที่ 4.3 จะเห็นได้ว่า อักษรติดกันมีผลต่ออัตราความถูกต้องน้อยมาก เพราะที่อักษร 14 - 24 point แม้ว่าจะมีอัตราอักษรติดกันต่างกันมากแต่มีอัตราความถูกต้องใกล้เคียงกัน แสดงว่าส่วนตัดแยกอักษรทำงานได้ดี ส่วนอักษรที่ขนาด 12 point มีอัตราความถูกต้องต่ำลงเป็นผลจากส่วนรู้จำอักษรเดี่ยวและอักษรติดกันที่เพิ่มขึ้นถึง 4 %



รูปที่ 4.4 แสดงอักษรแบบ JasmineUPC , KodchiangUPC เทียบกับ AngsanaUPC

จากรูปที่ 3.23 ในบทที่แล้วจะเห็นว่า BrowalliaUPC , DilleniaUPC , EucrosiaUPC , FreeciaUPC , IrisUPC มีลักษณะใกล้เคียง คือมีโครงสร้างหลักและลักษณะบ่งความต่างใกล้เคียงกับ AngsanaUPC และ CordiaUPC สำหรับ JasmineUPC , KodchiangUPC มีลักษณะต่างตั้งแต่โครงสร้างหลัก เช่น "ร" ของ JasmineUPC มีความกว้างจนไม่เป็นแบบ 1 ยอด หัวของ " ก ภ ถ " ของ JasmineUPC และหัวของ "น ม" ของ KodchiangUPC ไม่เด่นชัดเหมือน AngsanaUPC ช่องว่างด้านล่างของ " ท ห " ของ KodchiangUPC มีความใกล้เคียงกันจนใช้ลักษณะบ่งความต่างซึ่งใช้ได้ผลสำหรับ AngsanaUPC ใช้ไม่ได้

ตารางที่ 4.6 แสดงอัตราการรู้จำของเอกสารที่ได้รับการแก้ความเอียงแล้ว จะเห็นได้ว่าหลังแก้ความเอียงแล้วเอกสารยังคงเอียงอยู่แต่ไม่เกิน 1 องศาซึ่งทำให้ไม่มีผลกับส่วนตัดระดับของอักษรดังที่อธิบายในบทที่แล้ว แต่ผลความผิดพลาดเพิ่มขึ้นเมื่อองศาที่เอียงมากมีผลเนื่องจากสัญญาณรบกวนที่เกิดตอนหมุนเอกสารกลับ ดังรูปที่ 4.5

รูปที่ 4.5 แสดงสัญญาณรบกวนที่เกิดขึ้นเมื่อแก้เอกสารที่เอียง

เหตุการณ์นี้เกิดขึ้นเมื่อทำการหมุนเอกสารแล้วจุดที่หมุนมีเศษทศนิยมเมื่อปัดเศษแล้วจุดที่เคยอยู่ใกล้กันกลับไปทับกัน หรืออยู่ห่างกัน ซึ่งถ้าเอกสารเอียงไม่มากความผิดพลาดเกิดขึ้นน้อยกว่าดังรูปจะเห็นว่าที่แก้ความเอียงจาก 7 องศาจะมีสัญญาณรบกวนมากกว่าแก้ความเอียงจาก 5 องศา

เราสามารถวิเคราะห์ความผิดพลาดสามารถแยกตามการเกิดเป็น

3.1 การแบ่งคอลัมน์

3.2 การตัดบรรทัด

3.3 การตัดตัวอักษรที่ติด สามารถแยกได้อีกเป็น วิเคราะห์แบบที่ติดผิดและตัดผิดตำแหน่ง

3.4 การวิเคราะห์อักษรเดี่ยว

3.1 ความผิดพลาดจากการแบ่งคอลัมน์

ส่วนใหญ่ที่ผิดเกิดจากบางบรรทัดยาวกว่าบรรทัดอื่นมากทำให้ตัดอักษรของบรรทัดนั้นๆ หายไปซึ่งมีผลกับระบบโดยรวมน้อยมากเนื่องจากบทความในหนังสือปกติจะมีการตั้งกั้นหลังให้ตรงอยู่แล้ว สำหรับบทความที่ไม่ได้มีการตั้งกั้นหลังก็จะเกิดความผิดพลาดขึ้นได้

3.2 ความผิดพลาดจากการตัดแบ่งบรรทัด

เป็นต้นเหตุสำคัญของความผิดพลาดที่เกิดกับอักษรระดับบนและล่างซึ่งถ้าอักษรขนาดเล็กจะทำให้สระล่างลงมามีเส้นแนวระดับเท่ากับสระบนของอักษรบรรทัดล่าง เป็นผลทำให้ส่วนตัดบรรทัดทำงานผิดพลาดเมื่ออักษรใหญ่มาก ส่วนตัดบรรทัดจะเข้าใจผิดว่าช่องระหว่างอักษรบนกับกลางหรือกลางกับล่างเป็นเส้นแบ่งระหว่างบรรทัด (แม้ว่าเราทำการ paste (ซึ่งแสดงวิธีในบทที่ 3)ในแนวดิ่งแล้วก็ตาม ซึ่งจะไม่สามารถรวมระดับได้ถ้าช่องระหว่างระดับใหญ่มาก คือ อักษรที่ขนาดมากกว่า 24 point)

แนวทางแก้ไขต้องให้ผู้ใช้มีส่วนร่วมในการเลือกตัวแปรของระบบตัดบรรทัด (ซึ่งจะมีผลให้ไม่เป็นระบบอัตโนมัติ) คือ เมื่อผู้ใช้เห็นว่าอักษรมีขนาดใหญ่มากเกินขีดจำกัดของโปรแกรมต้องทำการบอกโปรแกรมให้เปลี่ยนตัวแปรของระบบเพื่อรับกับขนาดอักษรที่เปลี่ยนไป

3.3 ความผิดพลาดจากการตัดตัวอักษรที่ติด

จากการวิเคราะห์ตารางที่ 4.3 จะเห็นได้ว่าความผิดพลาดหลักๆ ไม่ได้เกิดจากส่วนตัดแยก แต่ส่วนตัดแยกก็ยังเป็นต้นเหตุของความผิดพลาดบางอย่าง

- การตัดอักษรติดแบบที่ 1 เช่น “วัติ” ไม้หันอากาศติดกับสระอิ ซึ่งซ้อนกันทำให้การตัดในแนวดิ่งทำให้อักษรทั้ง 2 ขาดส่วนสำคัญไป
- การตัดอักษรติดแบบที่ 2 เช่น “ ปั ปี ” จากบทที่แล้วเสนอวิธีตัดคือต้องลบหางของอักษระดับ 2 ออกก่อนแล้วจะเหลือแต่อักษรระดับบน ซึ่งหลายครั้งที่อักษรระดับบนถูกลบไปมากจนรู้จำผิด
- การตัดอักษรที่ติดกันแบบที่ 9 คือระดับกลางติดกันเอง เช่น “เว” เมื่อติดกันแล้วความกว้างไม่เกินเกณฑ์ที่กำหนดทำให้ถูกรู้จำเป็น “ก” เป็นต้น
- การตัดอักษรติดแบบ 3 4 5 6 7 และ 8 ส่วนใหญ่ถูกต้อง

3.4 ความผิดพลาดจากการรู้จำอักษรเดี่ยว

เมื่อความผิดพลาดหลักไม่ได้เกิดจากส่วนตัดแยกอักษรที่ติด เกิดเป็นบางครั้งจากการตัดคอลัมน์และบรรทัดผิด ดังนั้นการรู้จำอักษรเดี่ยวผิดจึงเป็นเหตุใหญ่ แต่ความผิดพลาดที่เกิดขึ้นมีค่าน้อยคือประมาณ 2 %

จากตารางที่ 4.2 เห็นว่าความผิดพลาดที่เกิดขึ้นเกิดกับอักษระดับบนและกลางพอๆ กันดังนั้นเมื่อหารกันแล้วเป็นเหตุของความผิดพลาดเฉลี่ยระดับละประมาณ 1%

สำหรับอักษรระดับบนความผิดพลาดเกิดขึ้นที่

- การตรวจสอบลักษณะปากนกของอักษร ซึ่งทำให้รู้จำผิด เช่น สระอิ เป็น ไม้ไต่คู้
- การตรวจสอบลักษณะบ่งความต่างที่ละเอียดอ่อนเช่น หางของสระ อี อื และ อื ซึ่งถูกรบกวนโดยสัญญาณรบกวนได้ง่าย

สำหรับอักษรระดับกลางความผิดพลาดเกิดขึ้นที่

- การหาโครงสร้างหลักผิด เช่น ตรวจ “ฉ” ได้ว่าไม่มีพื้นเรียบ
- หาลักษณะบ่งความต่างเฉพาะบางอย่างไม่พบ ซึ่งเป็นเหตุหลักของความผิดพลาด เนื่องจากเราพบว่าอักษรที่รู้จำผิดจะอยู่ในกลุ่มเดียวกัน เช่น

  ค -> ก , ด -> ค , ห -> ท , ญ -> ณ , ฐ -> ช , ว -> ง , ง -> เ , น -> บ , ม -> บ , ค -> ต , ค -> ศ , ค -> ฅ เป็นต้น

ซึ่งความผิดพลาดทั้งหมดที่กล่าวมาเป็นผลจากการสแกนทำให้เกิดสัญญาณรบกวนและผลจากการปรับแก้ความเอียง ซึ่งทำให้การหาลักษณะบ่งความต่างบางอย่างทำงานผิดไปเช่น สระ “า” เมื่อเอียงไปทางขวาส่วนล่างของอักษรจะยื่นออกมาเล็กน้อยซึ่งระบบจะเข้าใจผิดว่าเป็น “ว” เป็นต้น