การวิเคราะห์ผลทดสอบการฟังเสียง: การศึกษาเปรียบเทียบระหว่างการใช้คะแนนถูกผิดกับค่า d’

ธนศักดิ์ ศิริคะเณรัตน์[1]

(**Received:** *April 25, 2022;* **Revised:** *July 30, 2022;* **Accepted:** *August 29, 2022)*

บทคัดย่อ

ทฤษฎีตรวจจับสัญญาณ (Signal Detection Theory: SDT) เป็นทฤษฎีที่ใช้หลักการทางคณิตศาสตร์มาวิเคราะห์การตัดสินใจของมนุษย์ งานวิจัยด้านการฟังเสียงบางงานวิจัยได้นำแนวคิดของทฤษฎี SDT มาประยุกต์ใช้กับการวิเคราะห์ผลการทดสอบการฟัง อย่างไรก็ตาม จากการสำรวจงานวิจัยด้านการทดสอบการฟังเสียงพบว่า งานวิจัยบางงานเลือกใช้เพียงค่า d’ ซึ่งเป็นค่าที่คำนวณโดยอาศัยแนวคิดของ SDT และบางงานวิจัยใช้เพียงคะแนนถูกผิด จึงเกิดประเด็นคำถามว่า การใช้ค่า d’ กับคะแนนถูกผิด จะส่งผลต่อผลสรุปการวิเคราะห์การฟังหรือไม่ ผู้วิจัยได้ศึกษางานวิจัยที่ทดสอบการฟังเสียงจำนวน 7 งานวิจัยและนำผลคะแนนของทั้ง 7 งานวิจัยมาวิเคราะห์ค่า d’ และคะแนนถูกผิดเพื่อเปรียบเทียบว่าค่าทั้ง 2 ชนิดให้ผลที่แตกต่างกันหรือไม่ ผลการเปรียบเทียบพบว่าการใช้ค่าทั้ง 2 ชนิดนำไปสู่ข้อสรุปของการวิจัยที่ไม่แตกต่างกัน อย่างไรก็ตาม ข้อมูลจากการวิจัยครั้งนี้อาจยังไม่เพียงพอที่จะพิสูจน์ว่าการใช้ค่าทั้ง 2 ชนิดจะให้ผลลักษณะเดียวกันเสมอไป

คำสำคัญ: ทฤษฎีตรวจจับสัญญาณ, การฟังเสียง, การทดสอบการฟังเสียง, ค่าดีไพรม์

[1] คณะมนุษยศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ thanasak.si@ku.th

An analysis of speech perception test result: a comparison of using accuracy and d'

Thanasak Sirikenerat[2]

Abstract

Signal Detection Theory (SDT) uses mathematical methods to analyze human decision quality. Some research on speech perception applies SDT to analyze speech perception test results. More precisely, some speech perception researchers use only d' (a value calculated based on SDT) and some use only accuracy in their analysis. However, it is unknown if using only d' and only accuracy lead to different research results. In the present study, speech perception test data from five research studies were used to calculate d' and accuracy separately, and then compare results statistically. The data revealed that using either d' or accuracy do not make significant differences in terms of inferential statistical analysis. A study involving a larger pool of people studied would improve reliability.

Keywords: Signal Detection Theory, speech perception, perception test, d-prime

[2] Faculty of Humanities Kasetsart University thanasak.si@ku.th

บทนำ

ทฤษฎีตรวจจับสัญญาณ (Signal Detection Theory: SDT) เป็นแนวคิดทฤษฎีที่มีประวัติย้อนหลังไปถึงการทำงานของเรดาร์ตรวจจับเครื่องบินรบในสมัยสงครามโลกครั้งที่ 2 ในสมัยนั้น ผู้ควบคุมจอเรเดาร์จะต้องตัดสินว่าสัญญาณที่ตรวจจับได้นั้นเป็นเครื่องบินรบหรือไม่ จึงเกิดแนวคิดว่าเมื่อมนุษย์เป็นผู้ตัดสินย่อมต้องมีความแปรปรวนในการตัดสินใจ รวมทั้งเกณฑ์ในการตัดสินใจของแต่ละคนย่อมแตกต่างกันไป SDT จึงถือกำเนิดขึ้นเพื่อใช้หลักการทางคณิตศาสตร์ในการคำนวณคุณภาพในการตัดสินใจและเกณฑ์การตัดสินใจของคนแต่ละคน (Macmillan & Creelman, 2005) ต่อมาแนวคิดและวิธีการคำนวณใน SDT ได้ถูกนำมาประยุกต์ใช้กับสาขาอื่น ๆ เช่นการแพทย์ รังสีวิทยา นิติวิทยาศาสตร์ จิตวิทยา (Phillips et al., 2001) รวมถึงภาษาศาสตร์ด้วย (Iverson et al., 2003) โดยที่ทฤษฎีดังกล่าวเป็นที่ยอมรับและมีการประยุกต์ใช้ในสาขาต่างๆ มาตั้งแต่ช่วงปี ค.ศ. 1960-1970 (Pastore & Scheirer, 1974) ในกรณีภาษาศาสตร์ ทฤษฎี SDT มักพบใช้ในการศึกษาเกี่ยวกับการฟังเสียง (Huang & Ferreira, 2020)

งานวิจัยทางสัทศาสตร์และสัทวิทยาที่ศึกษาการฟังเสียง (speech perception) หลายงานวิจัยได้นำแนวคิดและหลักการคำนวณตามทฤษฎี SDT ไปใช้ในการวิเคราะห์ผลการทดสอบการฟัง จากการสำรวจงานวิจัยที่ศึกษาด้านการฟังเสียงในภาษาที่หนึ่ง ภาษาที่สองและงานวิจัยที่ศึกษาการฟังเสียงของผู้ที่มีความบกพร่องด้านการได้ยิน พบว่างานวิจัยแบ่งออกเป็น 2 กลุ่มคืองานวิจัยที่วิเคราะห์ผลด้วยทฤษฎี SDT เช่น งานวิจัยของ Iverson et al. (2003), Iversona and Kuhl (1995), Moseley et al. (2016), Rader et al. (2015) และงานวิจัยที่ไม่ใช้ทฤษฎี SDT ซึ่งวิเคราะห์ผลด้วยคะแนนถูกผิด (accuracy) เช่นงานวิจัยของ Christine (1998), Holderman (2015), Sirikanerat and Sugaya (2014), Sirikanerat and Sugaya (2016), Sirikanerat (2017) จึงนำไปสู่ประเด็นปัญหาว่า การวิเคราะห์ผลการทดสอบการฟังด้วยทฤษฎี SDT กับการวิเคราะห์ผลด้วยคะแนนถูกผิด จะได้ข้อสรุปที่แตกต่างกันหรือไม่

จากการศึกษางานวิจัยที่ผ่าน ๆ มาพบว่ายังไม่มีงานวิจัยใดที่ศึกษาเปรียบเทียบว่าการวิเคราะห์ผลการทดสอบการฟังโดยใช้ทฤษฎี SDT และไม่ใช้ทฤษฎี SDT จะก่อให้เกิดความแตกต่างในการตีความและการสรุปผลการวิจัยหรือไม่ ประเด็นปัญหาดังกล่าวอาจมีบทบาทสำคัญกับการวิจัยเกี่ยวกับการฟังเสียงเป็นอย่างยิ่ง หากการวิเคราะห์ผลการทดสอบการฟังเสียงโดยใช้วิธีที่แตกต่างกัน ให้ผลที่แตกต่างกัน ก็จำเป็นที่จะต้องศึกษาและวิเคราะห์ต่อไปว่าวิธีการใดเป็นวิธีการที่เหมาะสมกว่ากัน

งานวิจัยในครั้งนี้จึงมีวัตถุประสงค์ที่จะตอบคำถามดังกล่าวโดยการรวบรวมผลการศึกษาเกี่ยวกับการฟังจากงานวิจัยที่ผ่าน ๆ แล้ววิเคราะห์ผลของงานวิจัยเหล่านั้นโดยใช้ทฤษฎี SDT และไม่ใช้ทฤษฎี SDT จากนั้นจึงวิเคราะห์ทางสถิติเพื่อเปรียบเทียบว่าการวิเคราะห์ทั้ง 2 ลักษณะนำไปสู่ข้อสรุปเกี่ยวกับการฟังเสียงที่แตกต่างกันหรือไม่

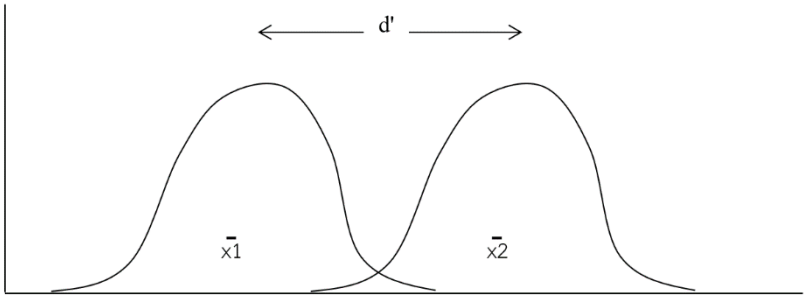
ทฤษฎี SDT

2.1 ภาพรวมของทฤษฎี SDT

ทฤษฎี SDT เริ่มนำมาใช้ครั้งแรกเพื่อการคำนวณคุณภาพการตัดสินใจของผู้ควบคุมเรดาร์ตรวจจับเครื่องบินรบในสมัยสงครามโลกครั้งที่ 2 โดยอาศัยหลักการทางคณิตศาสตร์และสถิติศาสตร์เป็นเครื่องมือสำคัญ (Pastore & Scheirer, 1974) โดยทฤษฎีนี้มีสมมติฐานว่ากระบวนการตัดสินใจของมนุษย์จะถูกรบกวนด้วยตัวแปรต่าง ๆ เช่น ลักษณะของสิ่งเร้า (stimuli), คุณภาพของสิ่งเร้า, เกณฑ์การตัดสินใจของมนุษย์ (response bias) ฯลฯ โดยทฤษฎี SDT มุ่งประเด็นที่ขจัดปัจจัยด้านเกณฑ์การตัดสินใจของมนุษย์ออกไปเพื่อให้ได้มากที่สุดโดยการใช้วิธีการทางคณิตศาสตร์

เมื่อเกณฑ์การตัดสินใจของมนุษย์แปรปรวนย่อมทำให้ผลที่ได้จากการเก็บข้อมูลไม่สามารถสะท้อนคุณภาพการตัดสินใจของมนุษย์ที่แท้จริงได้ (Macmillan, 2002; Macmillan & Creelman, 2005) ในกรณีของการทดสอบการฟัง การวิเคราะห์ผลด้วยคะแนนถูกผิดอาจจะไม่สะท้อนความสามารถในการฟังได้อย่างแท้จริง เนื่องจากคะแนนถูกผิดจะได้รับอิทธิผลจากเกณฑ์การตัดสินใจของผู้เข้าร่วมการทดลองเสมอ (Macmillan, 2002; Macmillan & Creelman, 2005) ทฤษฎี SDT จึงเสนอวิธีการวิเคราะห์คุณภาพการตัดสินใจที่ขจัดอิทธิพลของเกณฑ์การตัดสินใจออกเพื่อให้ได้ผลการทดลองที่สะท้อนคุณภาพการตัดสินใจที่แท้จริงให้ได้มากที่สุดโดยเสนอการใช้ค่าทางคณิตศาสตร์ซึ่งได้แก่ค่า d’ (d-prime)

โดยปกติ งานวิจัยที่ศึกษาการฟังเสียงมักมีวัตถุประสงค์เพื่อที่จะศึกษาว่าผู้เข้าร่วมการทดลองสามารถระบุหรือแยกเสียงได้หรือไม่ การวิเคราะห์ผลมักใช้คะแนนถูกผิดและหาค่าเฉลี่ยของคะแนนและตีความความสามารถในการฟังเสียงโดยอาศัยคะแนนถูกผิด ยกตัวอย่างเช่น หากเป็นการเทียบความสามารถในการฟังของผู้เข้าร่วมการทดลอง 2 กลุ่ม ก็จะได้ข้อมูลการกระจายของคะแนนเป็น 2 กลุ่มดังแสดงในรูปที่ 1 โดยที่เครื่องหมาย $\bar{x}$ แสดงค่าเฉลี่ย ตัวเลขหลังเครื่องหมาย $\bar{x}$ หมายถึงผู้เข้าร่วมการทดลองกลุ่มที่ 1 กับ 2 เมื่อได้ค่าเฉลี่ยและค่าทางคณิตศาสตร์อื่น ๆ เช่น ค่าเบี่ยงเบนมาตรฐาน (SD) ผู้วิจัยก็จะเปรียบเทียบค่าเฉลี่ยของผู้เข้าร่วมการทดลองทั้ง 2 กลุ่มเพื่อดูว่าค่าเฉลี่ยมีความแตกต่างกันหรือไม่ ซึ่งค่าเฉลี่ยในที่นี้ก็คือค่าเฉลี่ยของคะแนนถูกผิด ในขณะที่ SDT จะใช้ค่า d’ ซึ่งเป็นค่าระยะห่างระหว่าง $\bar{x}1$ กับ $\bar{x}2$ ตามที่แสดงในภาพที่ 1



ภาพที่ 1 ค่าเฉลี่ยและการกระจายตัวของข้อมูลสมมติ

โดยหลักการทางคณิตศาสตร์ได้พิสูจน์ให้เห็นแล้วว่าการใช้ค่า d’ ตามที่ทฤษฎี SDT เสนอนั้นจะมีค่าคงที่ ไม่แปรไปตามเกณฑ์การตัดสินใจของผู้เข้าร่วมการทดลอง ค่าดังกล่าวจึงน่าจะสะท้อนคุณภาพการตัดสินใจที่แท้จริงได้กว่าการใช้คะแนนถูกผิด (Macmillan, 2002; Macmillan & Creelman, 2005; McNicol, 1972; Pastore & Scheirer, 1974) อย่างไรก็ตาม แม้ตามหลักคณิตศาสตร์จะพิสูจน์ได้ว่าวิธีการคำนวณและค่าทางคณิตศาสตร์ที่ทฤษฎี SDT เสนอนั้นดูจะเป็นค่าที่บ่งบอกถึงคุณภาพการตัดสินใจที่ดี แต่ก็ยังคงเป็นที่น่าสงสัยว่าในทางปฏิบัติ การวิเคราะห์ผลจากการทดสอบการฟังในทางภาษาศาสตร์มีความจำเป็นที่จะต้องใช้วิธีการของ SDT หรือไม่

Phillips et al. (2001) ได้ศึกษาความแม่นยำในการพิสูจน์ลายเซ็นของหน่วยงาน FBI ในประเทศสหรัฐอเมริกา โดยเปรียบเทียบความแม่นยำในการตัดสินระหว่างเจ้าหน้าที่ FBI กับประชาชนทั่วไป ข้อมูลเบื้องต้นได้ผลออกมาว่าเจ้าหน้าที่ FBI สามารถตัดสินได้ว่าลายเซ็นที่ต้องสงสัยนั้นเป็นลายเซ็นเดียวกับที่พบในหลักฐานหรือไม่ถูกต้องประมาณ 50% ซึ่งใกล้เคียงกับประชาชนทั่วไปมาก Phillips et al. (2001) ชี้ให้เห็นว่าหากวิเคราะห์โดยอาศัยแต่เพียงคะแนนถูกผิดก็จะได้ข้อสรุปว่าเจ้าหน้าที่ FBI มีศักยภาพในการพิสูจน์ลายเซ็นไม่ต่างจากประชาชนทั่วไป แต่เมื่อวิเคราะห์ตามแนว SDT แล้วพบว่าเจ้าหน้าที่ FBI กับประชาชนทั่วไปมีเกณฑ์ในการตัดสินใจที่ต่างกันอย่างชัดเจน และเกณฑ์การตัดสินใจนี้ส่งผลกระทบต่อคะแนนถูกผิด

เมื่อขจัดปัจจัยด้านเกณฑ์การตัดสินใจแล้ววิเคราะห์ตามแนวทฤษฎี SDT กลับพบว่าเจ้าหน้าที่ FBI มีคุณภาพการตัดสินใจที่สูงถึง 90% ซึ่ง Phillips et al. (2001) ได้อภิปรายว่าปรากฏการณ์ดังกล่าวเกิดขึ้นเพราะว่าเจ้าหน้าที่ FBI มีเกณฑ์ในการตัดสินใจที่ระมัดระวังและเข้มงวดกว่าประชาชนทั่วไป จากงานวิจัยดังกล่าวเป็นการพิสูจน์ให้เห็นถึงประสิทธิภาพของการประยุกต์ใช้ทฤษฎี SDT อย่างชัดเจน แต่เป็นที่น่าเสียดายว่ายังไม่มีงานวิจัยที่พิสูจน์ว่าการใช้ทฤษฎี SDT ในการวิเคราะห์ผลการทดสอบการฟังจะได้ข้อสรุปที่แตกต่างจากการใช้คะแนนถูกผิดหรือไม่

2.2 ตัวอย่างการคำนวณค่า d’

เพื่อให้เห็นภาพที่ชัดเจนถึงวิธีการทางคณิตศาสตร์ที่ทฤษฎี SDT ใช้ ในที่นี้จะยกตัวอย่างการคำนวณค่า d’ สำหรับแบบทดสอบประเภท ใช่/ไม่ใช่ (yes/no response)[3] ซึ่งเป็นแบบทดสอบที่คำนวณค่า d’ ได้ง่ายที่สุด ยกตัวอย่างเช่น งานวิจัยสมมติชิ้นหนึ่ง[4] (ตัวอย่างสมมติปรับจาก Macmillan & Creelman (2005)) ต้องการทดสอบว่าผู้เข้าร่วมการทดลองจำได้หรือไม่ว่าเสียงพูดที่ตนได้ยินเป็นเสียงของคนที่เคยได้ยินมาก่อนแล้ว ในช่วงก่อนทดสอบ ผู้วิจัยให้ผู้เข้าร่วมการทดลองฟังเสียงพูดผู้ชาย 50 คน ช่วงทดสอบ ผู้เข้าร่วมการทดลองจะได้ฟังเสียงพูดของผู้ชาย 100 คน โดยแบ่งเป็นเสียงพูดของ 50 คนที่ผู้เข้าร่วมการทดลองได้ฟังไปก่อนแล้ว กับเสียงพูดของอีก 50 คนที่ผู้เข้าร่วมการทดลองไม่ได้ฟังในช่วงก่อนทดสอบ ผู้เข้าร่วมการทดลองต้องตอบว่าเสียงที่ได้ยินเป็นเสียงของ 1 ใน 50 คนที่ได้ฟังไปในช่วงก่อนทดสอบหรือไม่ หากผู้เข้าร่วม

[3] แบบทดสอบประเภท ใช่/ไม่ใช่ จะให้ผู้เข้าร่วมการทดลองดูหรือฟังสิ่งเร้าแล้วตอบว่าใช่หรือไม่ใช่

[4] ตัวอย่างสมมติใน Macmillan & Creelman (2005) เป็นตัวอย่างการทดสอบการรู้จำใบหน้า ผู้วิจัยได้ปรับตัวอย่างเป็นเรื่องการรู้จำเสียงเพื่อให้สอดคล้องกับวัตถุประสงค์งานวิจัยและเพื่อให้ผู้อ่านติดตามได้ง่ายขึ้น ลักษณะการทดลอง ตัวเลข ทั้งหมดเป็นเพียงข้อมูลสมมติเพื่อแสดงให้เห็นตัวอย่างการคำนวณเท่านั้น

การทดลองตอบว่า “ใช่” หมายความว่าผู้เข้าร่วมการทดลองตัดสินใจว่าเสียงที่ได้ยินเป็น เสียงพูดของ 1 ใน 50 คน ที่ได้ฟังไปก่อน ตาราง 1 คือข้อมูลสมมติจากงานวิจัยสมมติที่ได้อธิบายไป ตัวเลขในช่องแสดงจำนวนครั้งที่ตอบ “ใช่” หรือ “ไม่ใช่”

ตารางที่ 1 คือตารางเมทริกซ์ความสับสน (confusion metrix) หากผู้เข้าร่วมการทดลองได้ยิน “เสียงพูดที่ได้ฟังในช่วงก่อนทดสอบ” แล้วตอบว่า “ใช่” หมายความว่าตอบถูก (hit) หากตอบว่า “ไม่ใช่” หมายความว่าตอบผิด (miss) และหากผู้เข้าร่วมการทดลองได้ยิน “เสียงพูดที่ไม่ปรากฏในช่วงก่อนทดสอบ” แล้วตอบว่า “ใช่” หมายความว่าตอบผิด การตอบผิดในกรณีนี้เรียกว่า “สัญญาณหลอก” (false alarm) หากตอบว่า “ไม่ใช่” หมายความว่าตอบถูก การตอบถูกในกรณีนี้เรียกว่า “ปฏิเสธถูกต้อง” (correct negation)

ตารางที่ 1 ข้อมูลสมมติ

กลุ่ม 1			
สิ่งเร้า	คำตอบ		
	ใช่	ไม่ใช่	รวม
เสียงพูดที่ได้ฟังในช่วงก่อนทดสอบ	40 ถูก (Hit)	10 ผิด (Miss)	50
เสียงพูดที่ไม่ปรากฏในช่วงก่อนทดสอบ	20 สัญญาณหลอก (False alarm)	30 ปฏิเสธถูกต้อง (Correct negation)	50
กลุ่ม 2			
สิ่งเร้า	คำตอบ		
	ใช่	ไม่ใช่	รวม
เสียงพูดที่ได้ฟังในช่วงก่อนทดสอบ	16 ถูก (Hit)	34 ผิด (Miss)	50
เสียงพูดที่ไม่ปรากฏในช่วงก่อนทดสอบ	2 สัญญาณหลอก (False alarm)	48 ปฏิเสธถูกต้อง (Correct negation)	50

การคำนวณคะแนนถูกผิดจะต้องดูที่ความถี่ในการตอบถูกเท่านั้น ซึ่งได้แก่จำนวนครั้งที่ตอบในช่อง “ถูก” และ “ปฏิเสธถูกต้อง” ในตารางที่ 1 เมื่อเปรียบเทียบอัตราการตอบถูกของกลุ่มที่ 1 และ 2 โดยใช้คะแนนถูกผิดจะได้ผลสรุปว่า กลุ่มที่ 1 ตอบถูก 70% ในขณะที่กลุ่ม 2 ตอบถูก 62% หมายความว่ากลุ่ม 1 จดจำเสียงได้ดีกว่ากลุ่มที่ 2 อย่างไรก็ตาม หากเป็นการวิเคราะห์ตามแนวคิดทฤษฎี SDT จะใช้วิธีการคำนวณที่แตกต่างกันออกไป โดยขั้นต้น จะต้องคำนวณค่า H (“ถูก” Hit) และ F (“สัญญาณหลอก” False alarm) โดยมีวิธีคิดดังนี้

$$H = P(\text{"yes"}|S1)$$

$$F = P(\text{"yes"}|S2)$$

สูตรด้านบนสามารถอ่านได้ว่า ค่า H เท่ากับค่า P (สัดส่วน proportion) เมื่อตอบ “ใช่” ในกรณีที่ได้ยินหรือเห็นสิ่งเร้า (S = Stimuli) ที่ 1 และค่า F เท่ากับค่า P เมื่อตอบ “ใช่” ในกรณีที่ได้ยินหรือเห็นสิ่งเร้าที่ 2 จากข้อมูลสมมติในตารางที่ 1 สิ่งเร้าที่ 1 หมายถึง “เสียงพูดที่ได้ฟังในช่วงก่อนทดสอบ” สิ่งเร้าที่ 2 หมายถึง “เสียงพูดที่ไม่ปรากฏในช่วงก่อนทดสอบ” เมื่อคำนวณค่า P ตามสูตรด้านบนจะสามารถสรุปข้อมูลในตารางที่ 1 ใหม่ได้ดังนี้

ตารางที่ 2 ค่า P ของข้อมูลสมมติ

กลุ่ม 1			
สิ่งเร้า	คำตอบ		
	ใช่	ไม่ใช่	รวม
เสียงพูดที่ได้ฟังในช่วงก่อนทดสอบ	0.8 ถูก (Hit)	0.2 ผิด (Miss)	1.0
เสียงพูดที่ไม่ปรากฏในช่วงก่อนทดสอบ	0.4 สัญญาณหลอก (False alarm)	0.6 ปฏิเสธถูกต้อง (Correct negation)	1.0
กลุ่ม 2			
สิ่งเร้า	คำตอบ		
	ใช่	ไม่ใช่	รวม
เสียงพูดที่ได้ฟังในช่วงก่อนทดสอบ	0.32 ถูก (Hit)	0.68 ผิด (Miss)	1.0
เสียงพูดที่ไม่ปรากฏในช่วงก่อนทดสอบ	0.04 สัญญาณหลอก (False alarm)	0.96 ปฏิเสธถูกต้อง (Correct negation)	1.0

อันดับต่อมา เมื่อได้ค่า P แล้วก็จะต้องคำนวณหาค่า d’ ต่อไปโดยใช้สูตรต่อไปนี้

$$d' = z(H) - z(F)$$

ค่า z คือค่ามาตรฐาน (Standardized score) ที่ใช้ในสาขาคณิตศาสตร์ทั่วไป การหาค่า d’ สำหรับสูตรนี้จะต้องหาค่า z ของค่า H แล้วลบด้วยค่า z ของ F เนื่องจากการคำนวณที่ซับซ้อน Macmillan and Creelman (2005) จึงได้สรุปค่า z ของ H และ F เอาไว้เป็นตารางเพื่อความสะดวกในการวิเคราะห์ จากตารางค่า z ของ H และ F สามารถคำนวณหาค่า d’ ของกลุ่ม 1 และ 2 ได้ดังนี้

$$\text{กลุ่ม 1 } d' = 0.842 - (-0.253) = 1.095$$

$$\text{กลุ่ม 2 } d' = -0.468 - (-1.751) = 1.283$$

จากการคำนวณค่า d’ จะเห็นว่าคะแนนของกลุ่ม 2 ดีกว่ากลุ่ม 1 ซึ่งเป็นข้อสรุปที่แตกต่างจากการใช้คะแนนถูกผิดที่ได้แสดงไปตอนต้น (การใช้คะนนถูกผิดได้ข้อสรุปว่ากลุ่ม 1 คะแนนดีกว่า) จากวิธีการคำนวณที่ได้อธิบายไปข้างต้นจะเห็นว่าการคำนวณค่า d’ ตามแนวคิดทฤษฎี SDT มีข้อแตกต่างจากการใช้คะแนนถูกผิด

ทั่วไปก็คือ การคำนวณค่า d’ จะไม่ได้นับเฉพาะช่องที่ตอบถูกเท่านั้น แต่จะนับข้อที่ตอบผิดรวมเข้ามาอยู่ในการวิเคราะห์ด้วย จากตัวอย่างสมมติเบื้องต้นแสดงให้เห็นว่า การคำนวณผลการทดสอบการฟังโดยใช้คะแนนถูกผิดกับการใช้ทฤษฎี SDT อาจนำไปสู่ผลลัพธ์ที่แตกต่างกันได้ อย่างไรก็ตามตัวอย่างวิธีการคำนวณที่แสดงไปนั้นเป็นเพียงวิธีการคำนวณค่า d’ สำหรับแบบทดสอบประเภท “ใช่/ไม่ใช่” เท่านั้น หากเป็นแบบทดสอบประเภทอื่นก็จะใช้วิธีการคำนวณที่แตกต่างกัน

แบบทดสอบที่ใช้เพื่อทดสอบการฟังเสียง

ดังที่ได้กล่าวไปข้างต้นว่าหากใช้แบบทดสอบที่ต่างกัน วิธีการคำนวณค่า d’ ก็จะต่างกันด้วย การเลือกใช้แบบทดสอบ (task) ที่แตกต่างกันจะต้องใช้สูตรทางคณิตศาสตร์ที่แตกต่างกันในการคำนวณค่า d’ (Macmillan & Creelman, 2005; Pastore & Scheirer, 1974) เช่น แบบทดสอบ AX ใช้สูตร $d' = 2z\left(\frac{1}{2}\left\{1 + \left[2p(c)_{SD,IO} - 1\right]^{\frac{1}{2}}\right\}\right)$ โดยที่ p(c) หมายถึงสัดส่วนคำตอบที่ถูกต้อง แบบทดสอบ AFC ใช้สูตร $\mathrm{d}' = \frac{1}{\sqrt{2}}[\mathrm{z(H)} - \mathrm{z(F)}]$ โดยที่ H หมายถึงตอบถูก (Hit) และ F หมายถึงสัญญาณหลอก (False alarm) อย่างไรก็ตามจะเห็นได้ว่าไม่ว่าจะเป็นการทดสอบแบบใด ก็จำเป็นต้องใช้ผลการตอบถูกหรือผิดเป็นส่วนหนึ่งของการคำนวณค่า d’ ดังนั้นหากแบบทดสอบใดที่ไม่สามารถคำนวณสัดส่วนการตอบถูกผิดได้ก็จะไม่สามารถหาค่า d’ ได้ (Macmillan, 2002; Macmillan & Creelman, 2005)

Strange and Shafer (2008) ได้รวบรวมวิธีการออกแบบงานวิจัยและแบบทดสอบประเภทต่างๆ ที่มักใช้ในการทดสอบการฟังเสียงเอาไว้มากมาย แบบทดสอบการฟังหลักๆ ที่มักพบในงานวิจัยที่ทดสอบการฟังได้แก่ AX, ABX, Oddity, AFC (Alternative Forced Choice) และการให้คะแนนแบบสเกล

AX คือแบบทดสอบที่ใช้ไฟล์เสียง 2 ไฟล์ โดยให้ผู้เข้าร่วมการทดลองตอบว่าเสียงทั้ง 2 เสียงเหมือนกันหรือไม่ ในการทดสอบนี้เป้าหมาย (target) จะอยู่ที่สิ่งเร้า X ผู้เข้าร่วมการทดลองจะต้องตัดสินใจโดยยึดสิ่งเร้า X เป็นหลัก แล้วตัดสินใจว่าสิ่งเร้า X เหมือนหรือต่างกับสิ่งเร้า A แบบทดสอบนี้สามารถเปลี่ยนตำแหน่งของสิ่งเร้า X เช่น อาจเปลี่ยนเป็น XA โดยที่การเปลี่ยนตำแหน่งของสิ่งเร้า X ไม่มีผลต่อการคำนวณค่า d’

ABX คือแบบทดสอบที่ใช้ไฟล์เสียง 3 ไฟล์ เสียงเป้าหมายจะอยู่ที่สิ่งเร้า X โดยที่ผู้เข้าร่วมการทดลองต้องตอบว่าสิ่งเร้า X เหมือนสิ่งเร้า A หรือ B แบบทดสอบนี้สามารถเปลี่ยนตำแหน่งของสิ่งเร้า X ได้เช่นกัน เช่น XAB หรือ AXB เป็นต้น ขึ้นอยู่กับลักษณะการออกแบบของผู้วิจัยแต่ละคน โดยที่การเปลี่ยนตำแหน่งของสิ่งเร้า X ไม่มีผลต่อการคำนวณค่า d’

Oddity คือแบบทดสอบที่ใช้เสียง 3 เสียง โดยไม่มีสิ่งเร้าเป้าหมายเหมือนอย่างแบบทดสอบ 2 แบบที่ได้กล่าวไปข้างต้น แบบทดสอบนี้จะมี 1 เสียงที่มีลักษณะแตกต่างจากอีก 2 เสียง ผู้เข้าร่วมการทดลองจะต้องฟังเสียงทั้ง 3 เสียงแล้วตอบว่าเสียงข้อใดเป็นไฟล์เสียงที่แตกต่างจากอีก 2 ที่เหลือ นอกจากการใช้เสียง 3 เสียงแล้ว บางงายวิจัยอาจเพิ่มความยากด้วยการเพิ่มจำนวนเสียงให้มากกว่า 3 เสียงก็ได้ ประเด็นสำคัญของ

การวิเคราะห์แบบทดสอบ Oddity ในทฤษฎี SDT ก็คือจำนวนตัวเลือกจะส่งผลต่อค่า d’ แม้จะตอบถูกเหมือนกัน แต่หากจำนวนตัวเลือกมีมากกว่า ค่า d’ ก็จะสูงกว่า

AFC (Alternative Forced Choice) คือแบบทดสอบที่เล่นไฟล์เสียงเพียงเสียงเดียวแล้วให้ผู้เข้าร่วมการทดลองตอบว่าได้ยินเสียงอะไร จำนวนตัวเลือกอาจมีได้ตั้งแต่ 2 ตัวเลือก จนถึงเท่าใดก็ได้ ทั้งนี้จำนวนตัวเลือกจะขึ้นอยู่กับวัตถุประสงค์ของงานวิจัย ประเด็นสำคัญของการวิเคราะห์แบบทดสอบ AFC ในทฤษฎี SDT จะคล้ายกับกรณีแบบทดสอบ Oddity ก็คือ จำนวนตัวเลือกจะส่งผลต่อค่า d’ แม้จะตอบถูกเหมือนกัน แต่หากจำนวนตัวเลือกมีมากกว่า ค่า d’ ก็จะสูงกว่า

การให้คะแนนแบบสเกล เป็นแบบทดสอบที่ให้ผู้เข้าร่วมการทดลองฟังสิ่งเร้าแล้วให้เลือกตอบโดยเลือกค่าที่อยู่บนสเกล แบบทดสอบนี้สามารถปรับใช้ได้กับงานวิจัยหลากหลายลักษณะ เช่น การทดสอบว่าเสียงที่ฟังเป็นเสียงพูดของมนุษย์หรือไม่ การทดสอบว่าเสียงที่ฟังเป็นเสียงที่พูดโดยเจ้าของภาษาหรือไม่ การทดสอบว่าเสียงที่ฟังคล้ายเสียงที่มีในภาษาแม่ของผู้เข้าร่วมการทดลองหรือไม่ โดยปกติค่าบนสเกลมักจะกำหนดที่ 4-7 ค่า ประเด็นสำคัญของการวิเคราะห์การให้คะแนนแบบสเกลตามแนวทฤษฎี SDT คือ ผู้เข้าร่วมการทดลองต้องตอบ 2 ขั้นตอน โดยที่ขั้นตอนหนึ่งจะมีคำตอบที่ถูกต้อง ในขณะที่ขั้นตอนที่ 2 จะไม่มีคำตอบที่ถูกต้อง เช่น หากเป็นการทดสอบว่าเสียงที่ได้ยินเป็นเสียงที่พูดโดยเจ้าของภาษาหรือไม่ ผู้เข้าร่วมการทดลองจะต้องตัดสินใจ 2 ขั้นคือ 1 ตัดสินใจว่าเป็นเสียงของเจ้าของภาษาหรือไม่ ในการทดลองจะต้องมีคำตอบที่ถูกต้องสำหรับการตัดสินใจขั้นนี้อยู่ด้วย หมายความว่า ผู้เข้าร่วมการทดลองอาจจะตอบผิดหรือถูกก็ได้ และ 2 ตัดสินใจว่ามีความคล้ายคลึงเจ้าของภาษาเท่าใดโดยให้คะแนนบนสเกล การตัดสินใจนี้จะไม่มีคำตอบที่ถูกหรือผิด

ในการศึกษาครั้งนี้จะไม่ศึกษางานวิจัยที่ใช้การให้คะแนนแบบสเกลเนื่องจากว่าโดยปกติ งานวิจัยที่ศึกษาด้านการฟังเสียงจะให้ผู้เข้าร่วมการทดลองให้คะแนนบนสเกลเท่านั้นโดยไม่มีคำตอบที่ถูกต้อง ทำให้ไม่สามารถคำนวณค่า d’ ได้ การศึกษาครั้งนี้จึงจะรวบรวมผลของแบบทดสอบการฟังจากงานวิจัยที่ผ่านมาและวิเคราะห์ใหม่โดยเปรียบเทียบผลการวิเคราะห์ที่ใช้คะแนนถูกผิดและใช้ d’ เพื่อวิเคราะห์ว่าการวิเคราะห์ทั้ง 2 แบบจะนำไปสู่ข้อสรุปที่ต่างกันหรือไม่ ผู้วิจัยคัดเลือกเฉพาะงานวิจัยเกี่ยวกับการฟังเสียงที่ใช้แบบทดสอบประเภท AX, ABX, Oddity และ AFC เท่านั้น โดยงานวิจัยที่จะนำมาวิเคราะห์ผลนั้นจะต้องเป็นงานวิจัยที่มีข้อมูลคะแนนถูกผิดของผู้เข้าร่วมการทดลองรายบุคคลอยู่ด้วย สูตรในการคำนวณค่า d’ สำหรับทุกแบบทดสอบเป็นสูตรที่เสนอโดย Macmillan (2002); Macmillan and Creelman (2005)

วิธีการดำเนินการวิจัย

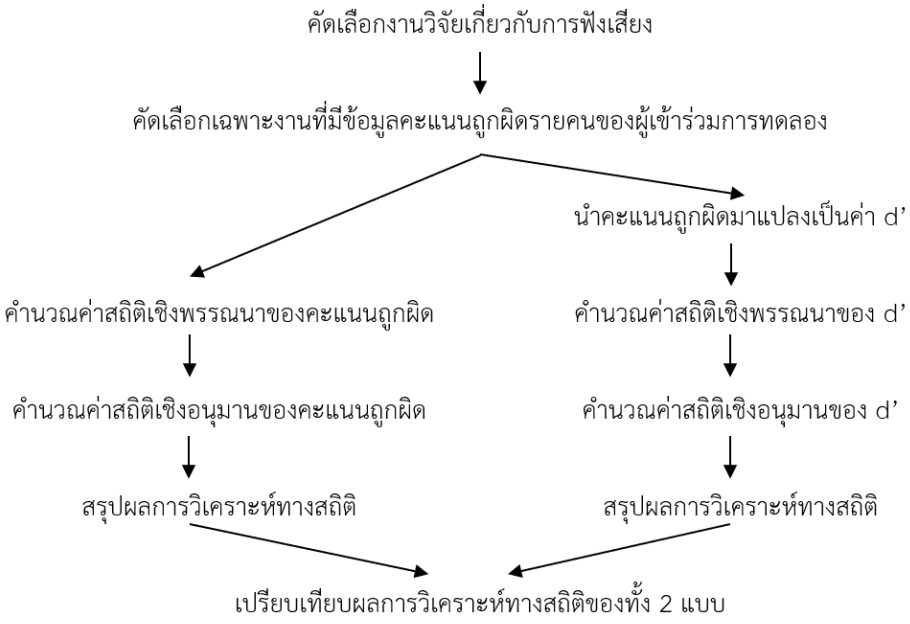
งานวิจัยนี้เป็นการนำข้อมูลงานวิจัยในอดีตที่ทดสอบการฟังมาวิเคราะห์ใหม่โดยใช้แนวคิดของทฤษฎี SDT และเปรียบเทียบผลการวิเคราะห์ทางสถิติเพื่อศึกษาว่าการวิเคราะห์ด้วยคะแนนถูกผิดกับการวิเคราะห์ด้วยค่า d’ จะพาไปสู่ข้อสรุปที่แตกต่างกันหรือไม่ โดยมีหลักเกณฑ์ในการคัดเลือกข้อมูลดังต่อไปนี้

1) ใช้ข้อมูลจากงานวิจัยทุกประเภทที่ทดสอบการฟัง ไม่ว่าจะเป็นงานวิจัยที่ตีพิมพ์ในวารสารหรือวิทยานิพนธ์ โดยไม่กำหนดว่าจะต้องเป็นการทดสอบการฟังประเภทใด การทดสอบการฟังอาจเป็นการทดสอบ

การฟังเสียงภาษาต่างประเทศ ภาษาแม่ หรือการฟังเสียงของผู้ที่มีความบกพร่องด้านการได้ยินก็ได้ เนื่องจากลักษณะของผู้ฟัง เสียงที่ฟัง และวัตถุประสงค์ของการวิจัยไม่มีผลต่อการวิเคราะห์ค่า d’

2) งานวิจัยที่จะนำมาวิเคราะห์จะต้องให้ข้อมูลคะแนนถูกผิดของผู้เข้าร่วมการทดลองรายคนไว้ด้วย เนื่องจากหากไม่มีข้อมูลรายคนจะทำให้ไม่สามารถคำนวณค่า d’ ได้ ดังนั้นงานวิจัยที่ให้ข้อมูลเพียงค่าเฉลี่ยจะไม่นำมาวิเคราะห์

ในการวิเคราะห์ ผู้วิจัยจะวิเคราะห์โดยนำคะแนนถูกผิดของแต่ละงานวิจัยมาวิเคราะห์ผลทางสถิติ จากนั้นจึงวิเคราะห์ผลอีกรอบโดยนำคะแนนถูกผิดมาคำนวณหาค่า d’ แล้ววิเคราะห์สถิติอีกครั้ง เมื่อได้ผลทางสถิติของคะแนนถูกผิดและ d’ แล้วจึงนำผลมาเปรียบเทียบกันว่าการวิเคราะห์ทั้ง 2 แบบนำไปสู่ข้อสรุปทางสถิติที่แตกต่างกันหรือไม่ ตามที่แสดงในภาพที่ 2



ภาพที่ 2 ขั้นตอนการดำเนินการวิจัย

ด้วยข้อจำกัดว่างานวิจัยที่จะนำมาวิเคราะห์เปรียบเทียบระหว่างคะแนนถูกผิดกับค่า d’ ได้ต้องมีคะแนนถูกผิดของรายบุคคลให้ไว้ด้วย ทำให้จำนวนงานวิจัยที่ใช้ได้มีน้อย จากการสืบค้นวิทยานิพนธ์และบทความวิจัยตั้งแต่ช่วงปี 2530-2564 พบงานวิจัยที่งานวิจัยที่ให้ข้อมูลคะแนนถูกผิดรายบุคคลไว้ด้วยเพียง 7 งานวิจัยเท่านั้น ซึ่งประกอบไปด้วยงานวิจัยที่ใช้แบบทดสอบ Oddity 3 งาน และ AFC 2 งาน และ AXB 1 งาน และ AX 1 งาน

ผลการวิเคราะห์

ในการนำเสนอผลการวิเคราะห์จะแยกการนำเสนอตามลักษณะของแบบทดสอบที่ใช้โดยไม่แยกวัตถุประสงค์การวิจัยหรือลักษณะของผู้เข้าร่วมการทดลอง เนื่องจากการวิธีการคำนวณค่า d’ จะขึ้นอยู่กับแบบทดสอบเท่านั้น ตารางที่ 3 แสดงข้อมูลงานวิจัยที่นำมาวิเคราะห์ในครั้งนี้

ตารางที่ 3 ค่าเฉลี่ยของคะแนนถูกผิดและค่าเฉลี่ย d'

งานวิจัย	แบบทดสอบ	วัตถุประสงค์	วิธีการทดสอบการฟัง	จำนวนผู้เข้าร่วมการทดลอง	ค่าเฉลี่ยคะแนนถูกผิด	ค่าเฉลี่ย d'
Sirikanerat and Sugaya (2014)	Oddity	ทดสอบการฟังเสียงสูงต่ำ (pitch accent) ในภาษาญี่ปุ่นของชาวต่างชาติ โดยเปรียบเทียบระหว่างกรณีที่สิ่งเร้าพูดโดยคนเดียวกับพูดโดย 3 คน	ฟังคำ 3 คำแล้วตอบว่าคำใดที่มีรูปแบบเสียงสูงต่ำแตกต่างจากอีก 2 คำที่เหลือ	8	**เสียงที่พูดโดย 1 คน** = 65.05% **เสียงที่พูดโดย 3 คน** = 53.47%	**เสียงที่พูดโดย 1 คน** = 2.60 **เสียงที่พูดโดย 3 คน** = 1.54
Sirikanerat and Sugaya (2016)				15	**เสียงที่พูดโดย 1 คน** = 67.20% **เสียงที่พูดโดย 3 คน** = 58.67%	**เสียงที่พูดโดย 1 คน** = 2.39 **เสียงที่พูดโดย 3 คน** = 1.92
Sirikanerat (2017)	Oddity	ทดสอบการฟังเสียงสูงต่ำ (pitch accent) ในภาษาญี่ปุ่นของผู้เรียนชาวจีน	ฟังคำ 3 คำแล้วตอบว่าคำใดที่มีรูปแบบเสียงสูงต่ำแตกต่างจากอีก 2 คำที่เหลือ	36	**คำที่แตกต่างคือคำที่ 1** = 81.64% **คำที่แตกต่างคือคำที่ 2** = 83.95% **คำที่แตกต่างคือคำที่ 3** = 83.49	**คำที่แตกต่างคือคำที่ 1** = 3.74 **คำที่แตกต่างคือคำที่ 2** = 3.90 **คำที่แตกต่างคือคำที่ 3** = 3.74
Christine (1998)	AFC	ทดสอบการฟังเสียง /s/ และ /ʃ/ ของผู้ที่มีความผิดปกติด้านการได้ยิน โดยวัดผล 3 ช่วงเวลาคือ 1	ฟังคำแล้วตอบว่าได้ยินเสียง /s/ หรือ /ʃ/	6	**เสียง /s/** 1 สัปดาห์ก่อนฝึก =52.78% 1 วันก่อนฝึก	**เสียง /s/** 1 สัปดาห์ก่อนฝึก =0.103 1 วันก่อนฝึก

งานวิจัย	แบบทดสอบ	วัตถุประสงค์	วิธีการทดสอบการฟัง	จำนวนผู้เข้าร่วมการทดลอง	ค่าเฉลี่ยคะแนนถูกผิด	ค่าเฉลี่ย d’
		สัปดาห์ก่อนฝึก 1 วันก่อนฝึก และหลังฝึก			=52.78% หลังฝึก =75.00% **เสียง /ʃ/** 1 สัปดาห์ก่อนฝึก =55.57% 1 วันก่อนฝึก =50.02% หลังฝึก =80.55%	=-0.117 หลังฝึก =1.205 **เสียง /ʃ/** 1 สัปดาห์ก่อนฝึก =0.207 1 วันก่อนฝึก =-0.018 หลังฝึก =1.430
Holderman (2015)	AFC	เปรียบเทียบการฟังเสียงที่มีเสียงรบกวนของแบบทดสอบ WIN และ S-WIN	ฟังแล้วเขียนคำที่ได้ยิน	22	**แบบทดสอบ WIN** =10.07 **แบบทดสอบ S-WIN** =6.91	**แบบทดสอบ WIN** =2.63 **แบบทดสอบ S-WIN** =2.32
Li (2015)	AXB	ฝึกฟังเสียง /θ/, /s/ และ /ð/, /z/ ในภาษาอังกฤษ	ฟังคำที่ขึ้นต้นด้วยเสียง /θ/, /s/ และ /ð/, /z/ แล้วตอบว่าคำที่เสียงพยัญชนะต้นเป็นเสียงเดียวกับคำที่ 2 คือคำที่ 1 หรือ 3	28	**คู่เสียง /θ/-/s/** ก่อนฝึก = 40.29% หลังฝึก = 92.13% **คู่เสียง /ð/-/z/** ก่อนฝึก = 39.85% หลังฝึก = 89.23%	**คู่เสียง /θ/-/s/** ก่อนฝึก = -0.876 หลังฝึก = 3.318 **คู่เสียง /ð/-/z/** ก่อนฝึก = -0.891 หลังฝึก = 2.805

งานวิจัย	แบบทดสอบ	วัตถุประสงค์	วิธีการทดสอบการฟัง	จำนวนผู้เข้าร่วมการทดลอง	ค่าเฉลี่ยคะแนนถูกผิด	ค่าเฉลี่ย d’
Phuangcharoen (1993)	AX	ทดสอบการฟังเสียงพยัญชนะ สระ และวรรณยุกต์ของผู้ที่สูญเสียการได้ยินย่านความถี่สูง	ฟังคู่ของคำที่ต่างเพียงแค่พยัญชนะ สระ หรือวรรณยุกต์แล้วตอบว่าคู่ดังกล่าวเสียงเหมือนกันหรือไม่	200	**เสียงพยัญชนะ** ผู้ที่ไม่มีความบกพร่องด้านการฟัง = 95.32% ผู้ที่สูญเสียการได้ยินย่านความถี่สูง = 87.35% **เสียงสระ** ที่ไม่มีความบกพร่องด้านการฟัง = 98.77% ผู้ที่สูญเสียการได้ยินย่านความถี่สูง = 94.62% **เสียงวรรณยุกต์** ที่ไม่มีความบกพร่องด้านการฟัง = 96.56% ผู้ที่สูญเสียการได้ยินย่านความถี่สูง = 92.44%	**เสียงพยัญชนะ** ผู้ที่ไม่มีความบกพร่องด้านการฟัง = 4.08 ผู้ที่สูญเสียการได้ยินย่านความถี่สูง = 3.27 **เสียงสระ** ที่ไม่มีความบกพร่องด้านการฟัง = 5.25 ผู้ที่สูญเสียการได้ยินย่านความถี่สูง = 4.68 **เสียงวรรณยุกต์** ที่ไม่มีความบกพร่องด้านการฟัง = 4.75 ผู้ที่สูญเสียการได้ยินย่านความถี่สูง = 4.12

ผลของค่าเฉลี่ยคะแนนถูกผิดกับค่าเฉลี่ยของ d’ ในตารางที่ 3 แสดงให้เห็นว่าโดยภาพรวมแล้ว ค่าเฉลี่ยของคะแนนถูกผิดจะสอดคล้องกับค่าเฉลี่ยของ d’ เช่น กรณีงานวิจัยของ Sirikanerat and Sugaya (2016) ค่าเฉลี่ยคะแนนถูกผิดของการฟังเสียงที่พูดโดย 1 คน สูงกว่าเสียงที่พูดโดย 2 คน ในทำนองเดียวกัน ค่าเฉลี่ย d’ ก็มีลักษณะเช่นเดียวกันคือค่าเฉลี่ย d’ ของการฟังเสียงที่พูดโดย 1 คน สูงกว่าเสียงที่พูดโดย 2 คน แต่ก็มีส่วนที่แตกต่างกันอยู่ ซึ่งปรากฏในกรณีงานวิจัยของ Christine (1998) ในส่วนค่าเฉลี่ยการฟังเสียง /s/ จากตารางที่ 3 จะเห็นว่าค่าเฉลี่ยคะแนนถูกผิดของการฟังเสียง /s/ ของช่วง 1 สัปดาห์ก่อนฝึกและ 1 วันก่อนฝึกมีค่าเท่ากัน คือ 52.78% ในคณะที่ค่าเฉลี่ย d’ กลับมีค่า 0.103 และ -0.117 ตามลำดับ และงานวิจัยของ Holderman (2015) ก็พบว่าค่าเฉลี่ยคะแนนถูกผิดของแบบทดสอบ WIN สูงกว่าแบบทดสอบ S-WIN อยู่ 31.4% แต่ค่าเฉลี่ย d’ กลับพบว่าค่าใกล้เคียงกันมาก โดยค่าเฉลี่ย d’ ของแบบทดสอบ WIN สูงกว่าเพียง 15.2%

อย่างไรก็ตามแม้ค่าเฉลี่ยส่วนหนึ่งในงานวิจัยของ Christine (1998) จะแตกต่างกัน และความห่างระหว่างคะแนนแบบทดสอบ WIN กับ S-WIN ในงานวิจัยของ Holderman (2015) จะลดน้อยลงเมื่อใช้ค่า d’ แต่ผลการวิเคราะห์ทางสถิติกลับพบว่าการใช้ค่าเฉลี่ยคะแนนถูกผิดกับค่าเฉลี่ย d’ พาไปสู่ข้อสรุปทางสถิติที่ไม่แตกต่างกัน ดังแสดงในตารางที่ 4

ตารางที่ 4 ผลการวิเคราะห์ทางสถิติ

งานวิจัย	ค่าเฉลี่ยคะแนนถูกผิด	ค่าเฉลี่ย d’	ผลสรุปการวิเคราะห์ทางสถิติ
Sirikanerat and Sugaya (2014)	Independent samples T-test t = -2.247 (28), *p = .033*	Independent samples T-test t = -2.230 (28), *p = .034*	เหมือนกัน
Sirikanerat and Sugaya (2016)	Mann Whitney U test U = 15.500, *p = .082*	Mann Whitney U test U = 15.000, *p = .073*	เหมือนกัน
Sirikanerat (2017)	Kruskal Walls test Kruskal-Wallis-H = .942, *p = .624*	Kruskal Walls test Kruskal-Wallis-H = .939, *p = .625*	เหมือนกัน
Christine (1998)	Friedman Test เสียง /s/ χ^2 = 2.472 (2), *p = .291* เสียง /ʃ/ χ^2 = 8.667 (2), *p = .0.13* Wilcoxon Signed Rank Test	Friedman Test เสียง /s/ χ^2 = 2.472 (2), *p = .291* เสียง /ʃ/ χ^2 = 8.667 (2), *p = .0.13* Wilcoxon Signed Rank Test	เหมือนกัน

<table>
<tr><th>งานวิจัย</th><th>ค่าเฉลี่ยคะแนนถูกผิด</th><th>ค่าเฉลี่ย d’</th><th>ผลสรุปการวิเคราะห์ทางสถิติ</th></tr>
<tr><td></td><td><table><tr><td></td><td>1w – 1d</td><td>post - 1w</td><td>post - 1d</td></tr><tr><td>z
p</td><td>-.736
.461</td><td>-2.041
.041</td><td>-2.214
.027</td></tr></table></td><td><table><tr><td></td><td>1w – 1d</td><td>post - 1w</td><td>post - 1d</td></tr><tr><td>z
p</td><td>-.736[b]
.461</td><td>-2.041
.041</td><td>-2.214
.027</td></tr></table></td><td></td></tr>
<tr><td>Holderman (2015)</td><td>Paired Samples T-test
t = 6.250 (21), p = .000</td><td>Paired Samples T-test
t = 11.905 (21), p = .000</td><td>เหมือนกัน</td></tr>
<tr><td>Li (2015)</td><td>Wilcoxon Signed Rank Test<table><tr><td></td><td>Pre-test vs Post-test /θ/-/s/</td><td>Pre-test vs Post-test /ð/-/z/</td></tr><tr><td>z
p</td><td>-4.543
.000</td><td>-4.543
.000</td></tr></table></td><td>Wilcoxon Signed Rank Test<table><tr><td></td><td>Pre-test vs Post-test /θ/-/s/</td><td>Pre-test vs Post-test /ð/-/z/</td></tr><tr><td>z
p</td><td>-4.542
.000</td><td>-4.542
.000</td></tr></table></td><td>เหมือนกัน</td></tr>
<tr><td>Phuangcharoen (1993)</td><td>เสียงพยัญชนะ
Mann Whitney U test
U = 1623.000, p = .000

เสียงสระ
Mann Whitney U test
U = 3929.000, p = .000

เสียงวรรณยุกต์
Mann Whitney U test
U = 3601.500, p = .000</td><td>เสียงพยัญชนะ
Mann Whitney U test
U = 1623.000, p = .000

เสียงสระ
Mann Whitney U test
U = 3929.000, p = .000

เสียงวรรณยุกต์
Mann Whitney U test
U = 3601.500, p = .000</td><td>เหมือนกัน</td></tr>
</table>

ตารางที่ 4 แสดงให้เห็นว่าแม้ค่าทางสถิติจะมีความแตกต่างบ้างเล็กน้อย (เช่นค่า t, U, z, χ^2) แต่ข้อสรุปจากการวิเคราะห์ทางสถิติไม่แตกต่างกัน อีกทั้งค่า *p* ในทุกงานวิจัยยังเท่ากัน หรือใกล้เคียงกันมากอีกด้วย

สรุปผลและอภิปราย

การวิจัยครั้งนี้ศึกษาเพื่อเปรียบเทียบว่า การใช้คะแนนถูกผิดกับการใช้ค่า d’ จะส่งผลให้ผลการวิเคราะห์ผลการทดสอบการฟังแตกต่างกันหรือไม่ จากผลการเปรียบเทียบระหว่างการใช้คะแนนถูกผิดกับค่า d’ พบว่ามีเพียงงานวิจัยเดียวเท่านั้นที่ค่าทางสถิติเชิงพรรณนาแตกต่างเล็กน้อย แต่โดยรวมแล้วค่าทั้ง 2 ก็มีลักษณะสอดคล้องกัน เมื่อวิเคราะห์ทางสถิติเปรียบเทียบแล้วพบว่าผลสรุปของการวิเคราะห์ทางสถิติไม่แตกต่างกัน กล่าวคือ การใช้จะใช้ค่าเฉลี่ยคะแนนถูกผิดกับ d’ ทำให้ได้ผลสรุปการวิจัยที่ไม่แตกต่างกัน

อาจเป็นไปได้ว่าการใช้ค่าเฉลี่ยคะแนนถูกผิดกับ d’ สำหรับการทดสอบ “ความสามารถในการฟัง” ของผู้เข้าร่วมการทดลองเท่านั้นอาจไม่แตกต่างกัน เช่น การทดสอบว่าผู้เข้าร่วมการทดลองฟังเสียงได้ถูกต้องหรือไม่ Huang and Ferreira (2020) ได้วิเคราะห์ผลการทดสอบการตัดสินความเป็นไปได้ทางไวยากรณ์ (acceptability judgement) ในภาษาอังกฤษและได้อภิปรายว่าการใช้ทฤษฎี SDT ทำให้วิเคราะห์ได้ว่าเกณฑ์การตัดสินใจของผู้เข้าร่วมการทดลองเปลี่ยนไปอย่างไรบ้างเมื่อสิ่งเร้าหรือสิ่งแวดล้อมเปลี่ยนไป จะเห็นว่าบทบาทของ SDT ที่ Huang and Ferreira (2020) ได้อภิปรายไว้นั้นไม่ใช่เรื่องความสามารถของผู้เข้าร่วมการทดลอง จากงานวิจัยของ Huang and Ferreira (2020) อาจเป็นไปได้ว่าหากใช้ทฤษฎี SDT เพียงเพื่อวิเคราะห์ว่าผู้เข้าร่วมการทดลองฟังเสียงได้ถูกต้องหรือไม่ ก็อาจไม่แตกต่างกับการใช้คะแนนถูกผิด

นอกจากประเด็นข้างต้นแล้ว ยังเป็นไปได้ว่าที่การวิจัยครั้งนี้ไม่พบความแตกต่างของการใช้และไม่ใช้ทฤษฎี SDT อาจเกิดขึ้นจากลักษณะการทดสอบในงานวิจัยที่เลือกมาวิเคราะห์ ผู้วิจัยได้กล่าวไปตอนต้นว่า Phillips et al. (2001) ได้ศึกษาความแม่นยำในการพิสูจน์ลายเซ็นของหน่วยงาน FBI และพบว่าการใช้เพียงคะแนนถูกผิดอาจไม่เหมาะสมกับการวิเคราะห์ แต่ในงานวิจัยครั้งนี้กลับพบว่า ไม่ว่าจะใช้ค่าประเภทใดก็ไม่ทำให้ข้อสรุปของงานวิจัยเปลี่ยนไป อย่างไรก็ตาม ในงานวิจัยของ Phillips et al. (2001) เก็บข้อมูลโดยใช้ผู้เข้าร่วมการทดลองตอบว่าลายเซ็นที่พบ เซ็นโดยคนเดียวกันกับในเอกสารหรือไม่ โดยให้ตอบว่าใช่หรือไม่ใช่ และยังให้ผู้เข้าร่วมการทดลองเลือกค่าความมั่นใจบนสเกลด้วยว่าผู้เข้าร่วมการทดลองมั่นใจกับคำตอบของตัวเองระดับใด การคำนวณผลในงานวิจัยของ Phillips et al. (2001) จึงต้องใช้ทั้งคะแนนถูกผิดและคะแนนความมั่นใจ ในขณะที่ข้อมูลจากงานวิจัยที่รวบรวมมาใช้ในการวิจัยครั้งนี้มีเพียงคะแนนถูกผิด เป็นไปได้ว่าหากเป็นงานวิจัยที่ทดสอบการฟัง และให้ผู้เข้าร่วมการทดลองเลือกระดับความมั่นใจในการตอบ อาจได้ผลออกมาลักษณะคล้ายกับงานวิจัยของ Phillips et al. (2001)

จากผลการวิเคราะห์ทางสถิติแสดงให้เห็นว่าการทดสอบการฟัง ไม่ว่าจะใช้คะแนนถูกผิดหรือใช้ค่า d’ ตามแนวคิดของทฤษฎี SDT ก็ไม่มีความแตกต่างกัน ซึ่งอาจตีความได้ว่าการทดสอบทักษะการฟัง สามารถวิเคราะห์ผลด้วยคะแนนถูกผิดได้โดยไม่จำเป็นต้องคำนวณค่า d’ ข้อดีของการใช้คะแนนถูกผิดก็คือสามารถคำนวณได้โดยง่ายดาย ต่างกับค่า d’ ที่จะเป็นต้องคำนวณด้วยสูตรทางคณิตศาสตร์ที่ซับซ้อน และยังต้องเลือกใช้สูตร

การคำนวณให้สอดคล้องกับลักษณะการเก็บข้อมูลอีกด้วย หรือกล่าวอีกนัยหนึ่งก็คือ การคำนวณค่า d’ สำหรับงานวิจัยที่ทดสอบทักษะการฟังอาจเป็นการเพิ่มขั้นตอนการวิเคราะห์ให้มากขึ้นเกินความจำเป็นก็เป็นได้

อย่างไรก็ตาม งานวิจัยครั้งนี้เปรียบเทียบผลการทดสอบการฟังจากงานวิจัยเพียง 7 งานเท่านั้น เนื่องจากข้อจำกัดที่ว่างานวิจัยส่วนใหญ่ไม่ได้รายงานคะแนนถูกผิดของผู้เข้าร่วมการทดลองรายคนไว้ ผู้วิจัยจึงไม่สามารถรวบรวมข้อมูลได้มากพอ ข้อมูลจากงานวิจัยเพียง 7 งานอาจยังไม่เพียงพอต่อการสรุปแน่ชัดว่าการใช้คะแนนถูกผิดกับค่า d’ ไม่มีความแตกต่างกัน หากวิเคราะห์งานวิจัยอื่นๆ เพิ่มขึ้นในอนาคตก็มีความเป็นไปได้ว่าอาจมีบางงานวิจัยที่จะได้ข้อสรุปที่แตกต่างกัน หากเป็นเช่นนั้นก็ต้องวิเคราะห์ต่อไปว่าสำหรับการทดสอบการฟังควรต้องใช้คะแนนถูกผิดหรือค่า d’ ในการสรุปผลการทดสอบ

นอกจากนี้ งานวิจัยนี้ยังมีข้อจำกัดอีกประการหนึ่งก็คือ ผู้วิจัยคำนวณค่า d’ จากคะแนนถูกผิด (accuracy) เท่านั้น การคำนวณค่า d’ สามารถทำได้ 2 ลักษณะคือ 1) คำนวณค่า d’ จากคะแนนถูกผิด และ 2 คำนวณค่า d’ จากคำตอบจริงของผู้เข้าร่วมการทดลอง การคำนวณแบบที่ 2 นี้ ต่างจากแบบแรกตรงที่การคำนวณแบบที่ 2 จะต้องคำนึงด้วยว่า ในข้อที่ตอบผิด ผู้เข้าร่วมการทดลองตอบว่าอะไร และนำคำตอบของผู้เข้าร่วมการทดลองมาแทนที่ค่าในสูตรการคำนวณด้วย ในขณะที่แบบที่ 1 จะใช้เพียงคะแนนรวมของแบบทดสอบในการคำนวณเท่านั้น ในการวิจัยครั้งนี้ผู้วิจัยเลือกใช้วิธีแบบที่ 1 เนื่องจากข้อมูลที่ผู้วิจัยใช้ในการวิจัยครั้งนี้มีเพียงคะแนนถูกผิดเท่านั้น ไม่มีข้อมูลว่าเมื่อผู้เข้าร่วมการทดลองตอบผิดอย่างไรบ้าง เป็นไปได้ว่าผลจากการวิจัยครั้งนี้ที่พบว่าการใช้คะแนนถูกผิดกับค่า d’ ไม่แตกต่างกัน อาจเนื่องมาจากการใช้วิธีการคำนวณแบบที่ 1 เพื่อศึกษาประเด็นดังกล่าว ควรมีการเปรียบเทียบการวิเคราะห์ผลด้วยคะแนนถูกผิดกับค่า d’ โดยใช้วิธีการคำนวณแบบที่ 2 เพื่อให้ได้ข้อสรุปที่ชัดเจนยิ่งขึ้น

กิตติกรรมประกาศ

บทความวิจัยนี้ได้รับทุนสนับสนุนจากทุนวิจัยโครงการภาคพิเศษ ภาควิชาภาษาศาสตร์ คณะมนุษยศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ ภายใต้โครงการ “บทบาทของทฤษฎีตรวจจับสัญญาณในการวิเคราะห์ผลการทดสอบการฟัง”

เอกสารอ้างอิง

Christine, K. (1998). *An investigation into the relatioship between speech perceptio and speech production in an experimental speech training program for children with hearing loss*. [Doctoral dissertation, The City University of New York].

Holderman, G. (2015). The accuracy of the Spanish Words in Noise (S-WIN) speech perception test in bilingual populations. [Bachelor of Arts thesis, University of Florida].

Huang, Y., & Ferreira, F. (2020). The application of Signal Detection Theory to acceptability judgments. *Frontiers in Psychology*, *31*, 1-11. https://doi.org/https://doi.org/10.3389/fpsyg.2020.00073

Iverson, P., Kuhl, P., K, Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87,* b47-b57.

Iversona, P., & Kuhl, P., K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of Acoustical Aociety of America*, *97*(1), 553-62.

Li, Y. (2015). Audio-visual training effect on L2 perception and production of English /θ/-/s/ and /ð/-/z/ by Mandarin speakers. [Doctoral, dissertation, Newcastal University].

Macmillan, N., A. (2002). Signal Dection Theory. In J. Wixed (Ed.), *Stevens' Handbook of Experimental Psychology* (3rd ed., Vol. 4). John Wiley & Sons, Inc.

Macmillan, N., A, & Creelman, C., Douglas. (2005). *Detection Theory: A User's Guide* (2nd ed.). Lawrence Erlbaum Associates Publishers.

McNicol, D. (1972). *A Primer of Signal Detection Theory*. Australasian Publishing.

Moseley, P., Smailes, D., Ellison, A., & Fernyhough, C. (2016). The effect of auditory verbal imagery on signal detection in hallucination-prone individuals. *Cognition*, *146*, 206–216.

Pastore, R., E, & Scheirer, C., J. (1974). Signal Detection Theory: considerations for general application. *Psychological Bulletin*, *81*(12), 945-958.

Phillips, V., L, Saks, M., J, & Peterson, J., L. (2001). The application of signal detection theory to decision-making in forensic science. *Journal of Forensic Sciences*, *46*(2), 294-308.

Phuangcharoen, W. (1993). *Speech perception ability of the persons who loss the high frequency.* [Doctoral dissertation, Mahidol University].

Rader, T., Adel, Y., Fastl, H., & Baumann, U. (2015). Speech perception with combined electric-acoustic stimulation: a simulation and model comparison. *Ear and hearing*, *36*(6), e314-325.

Sirikanerat, T. (2017). *The effect of F0 variation on the learning of Japanese pitch accent: a case study of Chinese speakers.* [Doctoral disseration, Tohoku University]. Tohoku University Repository. https://tohoku.repo.nii.ac.jp/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=124353&item_no=1&page_id=33&block_id=46.

Sirikanerat, T. & Sugaya, N. (2014). The effect of pitch variation in Japanese pitch accent on L2 learners: A pilot study. *Proceedings of the 28th General Meeting of The Phonetic Society of Japan*, Tokyo University of Agriculture and Technology Koganei Campus, Tokyo, 27-28 September (pp. 111-116).

Sirikanerat, T. & Sugaya, N. (2016). Difficulties in the perception of Japanese pitch accent by Thai learners using discrimination paradigm: The effects of accent type and multi-talker variation (in Japanese). *Bulletin of Institute for Excellence in Higher Education, Tohoku University*, 63-74.

Strange, W., & Shafer, V., L. (2008). Speech perception in second language learners: the re-education of selective perception. In J. G. H. Edwards & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition*. John Benjamins Publishing.