**การพูดแทรกเทคโนโลยีรู้จำเสียงพูด**

อีกด้านหนึ่งของเทคโนโลยีรู้จำเสียงพูดที่มีผลกระทบอย่างมากต่อการออกแบบก็คือการใช้งานของ Barge-In กล่าวคือจะอนุญาตให้ผู้ใช้ขัดจังหวะระบบเมื่อมีการพูดคุยหรือไม่ โดยทั่วไปแล้วระบบจะทำการเปิดใช้งาน Barge-in สำหรับการตอบกลับด้วยเสียง (IVR) เพื่อให้ผู้ใช้สามารถขัดจังหวะระบบได้ตลอดเวลา

เมื่อระบบตรวจพบคำพูดทุกชนิดมันจะหยุดเล่นทันทีที่และเริ่มฟัง เมื่อถึงจุดที่ผู้ใช้เริ่มพูด แต่ในขณะนั้นระบบก็ยังคงใช้คำสั่งต่อไป ตอนนี้ผู้ใช้คิดว่าได้ขัดจังหวะระบบก่อนที่จะเสร็จสิ้นและหยุดพูด แต่ก็สายเกินไป ระบบหยุดพูดเช่นกัน บทสนทนานั้นใช้งานไม่ได้และจะมีข้อความแจ้งข้อผิดพลาดเพื่อให้ผู้ใช้กลับมาติดตาม ลองนึกภาพการพูดคุยกับใครบางคนในการเชื่อมต่อโทรศัพท์มือถือที่ไม่ได้

อาจมีความล่าช้าที่สังเกตเห็นได้ซึ่งไม่มีอยู่ในคนและผู้โทรมักพูดคุยกัน Barge-in ยังมีประโยชน์อย่างมากเมื่อระบบทำการดำเนินการที่อาจใช้เวลานานหรืออ่านข้อมูลจำนวนมาก ตัวอย่างเช่นเมื่อ Amazon Echo กำลังเล่นเพลงผู้ใช้สามารถแทรกขึ้นมา เวลาใดก็ได้และพูดว่า “Alexa หยุด” หากไม่มีการแทรกแซง จะไม่มีทางหยุดการเล่นเพลงโดยใช้คำสั่งเสียง ซึ่งแตกต่างจากระบบ IVR แบบดั้งเดิม Alexa ไม่หยุดพูดเมื่อตรวจพบคำพูดใด ๆ

เฉพาะคำปลุกของเธอเท่านั้นสิ่งนี้เรียกว่าคำว่า hotword หรือ magic มันเป็นระบบที่ซับซ้อนมากเพราะมันไม่ได้หยุดระบบจากการแสดงการพูด เว้นแต่จะมีการจดจำคำหลักหรือวลีเฉพาะ สิ่งนี้สำคัญมากในบางสถานการณ์ ลองจินตนาการว่าขอให้ Alexa เล่นสถานีวิทยุ Pandora โดยเฉพาะ ในขณะเดียวกันผู้ใช้เริ่มพูดคุยกับครอบครัวของคุณ การที่ Alexa พูดว่า “ขออภัยนั่นคืออะไร” หลังจากที่ผู้ใช้พูดจะเป็นประสบการณ์ที่แย่มากสำหรับผู้ใช้ ระบบเพิกเฉยต่อความสุขของผู้ใช้จนกว่าคำพูดปลุกของเธอจะได้รับการยอมรับอย่างมั่นใจ น่าจะเป็นทางเลือกที่ดีกว่า Hotwords นั้นถูกใช้ในโลก IVR เช่นกัน แต่ในบริบทที่เฉพาะเจาะจง ตัวอย่างหนึ่งคือระบบ IVR ของ San Francisco Bay Area 511 ผู้ใช้สามารถโทรเพื่อรับข้อมูลการจราจรและเวลาในการขับขี่โดยประมาณ หลังจากให้ชื่อทางหลวงระบบจะค้นหาเหตุการณ์ที่เกี่ยวข้องและเริ่มอ่านพวกเขากลับไปที่ผู้โทรระบบต้องการให้ผู้ใช้มีความยืดหยุ่นในการข้ามไปยังเหตุการณ์ถัดไป แต่กังวลเกี่ยวกับเสียงรบกวนรอบด้านในรถยนต์หรืออินพุตอื่นรบกวนระบบและหยุดทำงาน ลองนึกภาพว่าผู้ใช้กำลังฟังรายการอุบัติเหตุ 10 รายการและผู้ใช้จาม ทำให้ระบบหยุดและพูดว่า “ฉันขอโทษฉันไม่ได้รับสิ่งนั้น” แน่นอนว่าเหตุการณ์แบบนี้จะทำให้เกิดประสบการณ์ที่ไม่ดีแน่ๆ วิธีแก้ปัญหานี้อาจจะใช้ การจดจำวลีสำคัญเพียงไม่กี่คำในระหว่างการอ่านข้อมูลนี้เช่น “ถัดไป” และ “ก่อนหน้า” เมื่อผู้ใช้พูดระบบจะไม่ตัดพรอมต์ทันทีเช่นเดียวกับ Barge-In อื่นๆ ในโหมด แต่ดำเนินการต่อเว้นแต่จะมีการจดจำคำหลักคำใดคำหนึ่งแล้วหยุดและย้ายไปยังการทำงานถัดไป อีกตัวอย่างหนึ่งที่ใช้ คือเมื่อผู้ใช้ต้องการหยุดชั่วคราวเพื่อดำเนินการให้เสร็จสิ้นในระหว่างการสนทนา สิ่งนี้อาจเกิดขึ้นได้หากผู้ใช้จำเป็นต้องไปและหยิบของบางอย่างเพื่อตอบคำถามเช่นรับขวดยาเพื่อตรวจสอบหมายเลขใบสั่งยาในระหว่างที่ขอให้เติมเงิน ระบบถามว่า

“คุณต้องการเวลาในการค้นหาหมายเลขใบสั่งยาหรือไม่” และหากผู้ใช้แจ้งว่า “ใช่” ระบบจะสั่งให้ผู้ใช้พูดว่า “ฉันกลับมาแล้ว” หรือ “ทำต่อ” เมื่อพร้อม หยุดการสนทนาชั่วคราว มันเป็นไปได้ที่จะปรับความไวของมันด้วยเครื่องมือ ASR บางตัว โดยพื้นฐานแล้วนักพัฒนาสามารถทำให้มีความไวมากขึ้นหรือน้อยลง (ยิ่งมีความละเอียดอ่อนน้อยลงเท่าใดก็จะยิ่งยากขึ้นสำหรับผู้ใช้ในการ จำกัด ความไว) บ้างเครื่องมือ ASR สามารถใช้ตรวจหาการจบประโยค ซึ่งหมายถึงการหยุดชั่วคราวในสิ่งขณะที่ผู้ใช้พูดก่อนที่ระบบจะตัดสินว่าผู้ใช้พูดเสร็จแล้ว เอ็นจิ้นรู้จำเสียงบางตัว ไม่สามารถตั้งค่าการหมดเวลาการพูดเมื่อสิ้นสุดการพูด แต่ใช้วิธี การหยุด 1.5 วินาทีสำหรับการตอบสนองหลัง ผู้ใช้พูดประโยคสุดท้าย ซึ่งมันอาจยาวเกินไปและทำให้ผู้ใช้จะสงสัยว่าระบบได้ยินหรือไม่ ระบบ VUI ที่ออกแบบมาอย่างดีที่สุดนั้นมีความยืดหยุ่นเพียงพอที่จะมีค่าการหมดเวลาที่แตกต่างกันในแต่ละสถานะ ตัวอย่างเช่นการโต้ตอบที่ผู้ใช้เป็นผู้เริ่มต้น (เช่นพูดว่า “ตกลง Google” หรือกดปุ่มโฮมของ Apple เพื่อเปิดใช้งาน Siri) ต้องใช้เวลาที่สั้นกว่าการตอบสนองต่อ “คุณเป็นอย่างไรวันนี้” ในกรณีแรก ผู้ใช้เริ่มต้นเหตุการณ์ไม่ใช่ระบบมีโอกาสที่ผู้ใช้จะรู้ว่าสิ่งที่พวกเขากำลังจะพูดและไม่จำเป็นต้องหยุดชั่วคราวนาน ในกรณีที่สองผู้ใช้อาจหยุดและเริ่มต้นเล็กน้อย ตัวอย่างเช่น “ฉันรู้สึก…ดีก่อนหน้านี้ฉันก็โอเค แต่ตอนนี้ฉัน…หัวฉันกำลังเจ็บ” ในกรณีนี้ถ้าการหมดเวลาสั้นเกินไปผู้ใช้จะถูกตัดออกก่อน เสร็จสิ้นซึ่งเป็นการสนทนาที่หยาบคายมาก อีกกรณีทั่วไปที่ต้องใช้เวลาหมดเวลาในการพูดอีกต่อไปคือเมื่อผู้คนอ่านหมายเลขที่จัดกลุ่มตามธรรมชาติเช่นบัตรเครดิต ผู้คนจะหยุดระหว่างกลุ่มโดยธรรมชาติและระบบต้องไม่มีการตัดการสนทนา วิธีที่ดีที่สุดที่จะรู้วิธีปรับแต่งคือการใช้ข้อมูล เมื่อดูที่การถอดความจากสิ่งที่ผู้คนพูดจริง ๆ นักพัฒนาสามารถค้นหาบริบทที่ผู้ใช้มักถูกตัดการสนทนากลางประโยค ในกรณีนี้คุณจะต้องทดลองขยายเวลาหมดเวลาการพูด การขยายระยะเวลาจบประโยค มีประโยชน์คือเมื่อคาดหวังให้ผู้ใช้พูดมากหรือลังเล ตัวอย่างเช่นการขอให้ผู้ใช้เล่ารายละเอียดของอุบัติเหตุทางรถยนต์สำหรับแอปประกัน ผู้ใช้มีแนวโน้มที่จะพูดหลายประโยคหยุดชั่วคราวในขณะที่รวบรวมความคิดของพวกเขา ในบางกรณีควรลดระยะเวลาสิ้นสุดการพูดให้สั้นลง เมื่อผู้ใช้เพียงแค่พูดว่า “ใช่” หรือ “ไม่” การหมดเวลาที่สั้นลงอาจนำไปสู่บทสนทนา ที่ตอบสนองได้ดีกว่า

การหมดเวลาจบประโยคที่สำคัญอีกข้อหนึ่งคือการไม่พบคำพูด (No Speech Detected :NSP) ควรถือว่าเป็นการหมดเวลาแยกต่างหากจากการพูดจบด้วยเหตุผลหลายประการคือ

NSP ยาวกว่าการหมดเวลาจบประโยคเมื่อพูดจบ (ปกติประมาณ 10 วินาที)

NSP ส่งผลให้เกิดการกระทำที่แตกต่างกันโดยระบบ VUI

มีประโยชน์ในการวิเคราะห์ระบบเพื่อพิจารณาปัญหา

ในระบบ IVR การหมดเวลาสนทนาแบบ NSP เกิดขึ้นเมื่อระบบเริ่มฟังการตอบสนองของผู้ใช้และไม่ตรวจพบคำพูดใด ๆ ในช่วงระยะเวลาหนึ่ง จากนั้นขึ้นอยู่กับนักออกแบบ VUI ที่จะตัดสินใจว่าจะทำอย่างไรในกรณีนี้ ด้วยระบบ IVR ผู้ใช้มักจะได้รับข้อความแสดงข้อผิดพลาดเช่น“ ขออภัยฉันไม่ได้ยิน คุณกำลังเดินทางวันใด” และรอให้ผู้ใช้พูด บางระบบไม่ทำอะไรเลยเมื่อหมดเวลาทำงานของ NSP ตัวอย่างเช่นหากพูดว่า “Alexa”

เพื่อเปิดใช้งาน Amazon Echo จากนั้นไม่พูดอะไรอีกหลังจากนั้นประมาณแปดวินาทีแสงสีฟ้าที่ด้านบนของอุปกรณ์จะปิดลงและ Alexa จะนิ่งเงียบ ถ้าเป็น google assistant หากพูดว่า “OK Google “ รอประมาณห้าวินาทีและหากไม่มีการพูดถึง ระบบจะแสดงหน้าจอพร้อมตัวอย่างสิ่งที่ผู้ใช้สามารถพูดได้ การไม่ตอบสนองอะไรเลยเป็นบางครั้งเป็นทางเลือกที่ดีกว่า ในกรณีเหล่านี้ผู้ใช้ที่ระบบไม่ได้ยินก็ชัดเจนและไม่ตอบสนองเลย จำทำให้ผู้ใช้ลองอีกครั้ง ในกรณีที่ระบบไม่แสดงผลผู้ใช้อย่างชัดเจนเกี่ยวกับการหมดเวลา NSP — แสดงให้เห็นถึงโหมดต่างๆของ VUIs ที่ใช้กันทั่วไปในปัจจุบัน ผู้ช่วยเสมือนเหล่านี้จำนวนมากยังคงอยู่ในโหมด“ one-off” กล่าว ระบบคาดหวังให้ผู้ใช้พูดอะไรบางอย่างและตอบกลับและโดยทั่วไปการสนทนาจะจบลงจนกว่าผู้ใช้จะเริ่มต้นคำขอใหม่ ในโลก IVR ผู้ใช้อยู่กลางกล่องโต้ตอบและจะไม่สามารถก้าวไปข้างหน้าได้หากไม่ได้รับการป้อนข้อมูลมากขึ้นดังนั้นการเตือนผู้ใช้ในการหมดเวลาของ NSP ก็อาจจะสมเหตุสมผลมากขึ้น อย่างไรก็ตามเป็นสิ่งสำคัญที่ต้องทำเพิ่มเติมสำหรับเหตุการณ์ NSP เมื่อผู้ใช้ติดขัด หากอยู่ในระบบการสนทนาที่มีอวาตาร์และการหมดเวลา NSP หลายครั้งได้รับการกระตุ้นให้ผู้ใช้หมดเวลา หากระบบมีส่วนต่อประสานกับผู้ใช้แบบกราฟิก (GUI) แสดงอยู่ (เช่นปุ่มบนหน้าจอ) แสดงว่าเพียงพอแล้ว GUIs สามารถรอจนกว่าผู้ใช้จะดำเนินการ ในขณะที่ออกแบบระบบ ให้ใช้เวลาคิดดูว่าเหตุใดการหมดเวลาการจบประโยค ของ NSP จึงถูกเรียกขึ้นมา ครั้งแรกแม้ว่าระบบคิดว่ามันไม่ได้ยินคำพูดใด ๆ แต่ก็อาจเป็นไปได้ที่ผู้ใช้พูด แต่ไม่มีระบบไม่สามารถรับเสียงได้

นักออกแบบบางคนสร้างบทที่พูดสิ่งต่าง ๆ เช่น“ พูดดังขึ้น” หรือ“ เข้าใกล้ไมโครโฟนมากขึ้น” แต่สิ่งเหล่านี้อาจทำให้ผู้ใช้หงุดหงิดมากถ้าผู้ใช้พูดเบาเกินไปที่จะได้ยินให้บอกพวกเขาให้พูดดัง ๆ วิธีนี่ไม่ค่อยแก้ปัญหา แทนที่จะออกแบบวิธีการนำผู้ใช้ไปยังขั้นตอนถัดไปแทน บ่อยครั้งซึ่งหมายความว่าการอนุญาตให้ผู้ใช้ทำซ้ำตัวเองหรือหลังจากหมดเวลา NSP หลายครั้งเป็นการให้ทางเลือกอื่นในการป้อนข้อมูล มีบางครั้งที่ผู้ใช้ไม่ได้พูดอะไรเลย ในฐานะนักออกแบบ VUI ควร พิจารณาว่าทำไมสิ่งนี้อาจเป็น หากข้อมูล แสดงให้เห็นถึง สถานที่เฉพาะในแอปที่ผู้ใช้ไม่พูดให้ตรวจสอบการโต้ตอบอย่างใกล้ชิดยิ่งขึ้น นี่คือตัวอย่างของแอปที่อนุญาตให้ผู้ใช้ชำระค่าบริการของผู้ให้บริการอินเทอร์เน็ต (ISP) ด้วยคำถามที่นำไปสู่การหมดเวลา NSP บ่อยครั้ง ตัวอย่างต่อไปนี้แสดงวิธีจัดการอย่างไม่ดี

ISP VUI: หมายเลขบัญชีของคุณคืออะไร

ผู้ใช้ : [ความเงียบ]

ISP VUI: ฉันขอโทษฉันไม่ได้ยินอะไรเลย กรุณาพูดหมายเลขบัญชีของคุณ

ผู้ใช้: [ความเงียบ]

ISP VUI: ฉันยังไม่ได้ยินอะไรเลย กรุณาพูดหมายเลขบัญชีของคุณ

จะเห็นได้ว่า วิธีการนี้ไม่ได้นำไปสู่ผลลัพธ์ที่ประสบความสำเร็จ เหตุผลก็เพราะผู้ใช้ไม่ได้รับความช่วยเหลือ ระบบจะทำซ้ำคำถาม คำถามลักษณะนี้นำไปสู่การหมดเวลาจบประโยค NSP จำนวนมากบ้างที ผู้ใช้พยายามชำระยอดเงินคงเหลือ แต่ไม่ทราบหมายเลขบัญชี อาจจะดีกว่าถ้านักออกแบบ ออกแบบให้ตอบโต้ตามตัวอย่างต่อไปนี้

ISP VUI: หมายเลขบัญชีของคุณคืออะไร

ผู้ใช้: [ความเงียบ]

ISP VUI: ขออภัยฉันไม่เข้าใจ หมายเลขบัญชีของคุณสามารถพบได้ที่ด้านบนของใบแจ้งยอดของคุณ โปรดพูดหรือพิมพ์ลงในหรือพูดว่า “ฉันไม่รู้”

ผู้ใช้: ฉันไม่รู้

ISP VUI: ไม่มีปัญหา. เราสามารถค้นหาด้วยหมายเลขโทรศัพท์และที่อยู่ของคุณแทน …

การประโยคอีกลักษณะหนึ่ง คือการพูดมากเกินไป (Too-Much-Speech ,TMS) ซึ่งจะถูกเรียกใช้ในกรณีที่ผู้ใช้กำลังพูดเป็นเวลานานโดยไม่มีการหยุดชั่วคราวซึ่งโดยปกติจะเรียกใช้การหมดเวลาเมื่อสิ้นสุดการพูด สำหรับระบบส่วนใหญ่โดยทั่วไปไม่จำเป็นต้องจัดการกรณีนี้เนื่องจากผู้ใช้จะต้องหายใจในบางจุด เป็นความคิดที่ดีที่จะยังคงมองหากิจกรรมเหล่านี้ในแอปพลิเคชันที่ปรับใช้เพราะอาจบ่งชี้ว่าตัวจำแนกลายมือกำลังเรียกใช้คำพูดที่ไม่มีความหมายและนักพัฒนาต้องระบุสาเหตุ อย่างไรก็ตามหากพบว่าตัวเองออกแบบระบบที่กระตุ้นให้ผู้ใช้พูดเป็นส่วนที่มีความยาวและความยาวของคำพูดยาวเกินไปนักพัฒนา สามารถกำหนดเวลาการจบประโยคของ TMS และตัดผู้ใช้ออกเพื่อสนทนาต่อไป ควรดูข้อมูลเพื่อกำหนดเวลาหมดของ TMS แต่สามารถเริ่มต้นด้วยสิ่งที่ไม่สั้นเกินไป (และจะตัดผู้ใช้ออกได้ง่ายเกินไป) เช่น 7 ถึง 10 วินาที