

Agency in Augmentative and Alternative Communication (AAC)

Stephanie Valencia Valencia

CMU-HCII-23-104

July 2023

Human-Computer Interaction Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA, USA

Dissertation Committee:

Henny Admoni (Co-chair), Carnegie Mellon University
Jeffrey P. Bigham (Co-chair), Carnegie Mellon University
Jodi Forlizzi, Carnegie Mellon University
Jeff Higginbotham, University at Buffalo, SUNY

*A dissertation submitted in partial fulfillment
of the requirements for the degree of Doctor of Philosophy in
Human-Computer Interaction at Carnegie Mellon University*

*“Accessibility is a continuous and evolving practice. Accessibility is achieved through intentional, meaningful and intersectional participation of people with lived experience of exclusion. Accessibility must be key to each stage of a product, place or policy development, from ideation through to delivery. Solutions must be designed with disabled people to prioritise form and function. Meaningful and deliberate accessibility builds inclusion, equity, **agency**, creativity, innovation and pride.”*

Tilting the Lens’ working definition of accessibility.

Tilting the Lens is led by Sinéad Burke, Irish writer, academic and disability activist. This definition reflects my current views and approaches towards designing for agency in AAC.

Abstract

Augmentative and alternative communication (AAC) devices enable speech-based communication when people cannot use their own physical voice to speak. Speech-generating AAC devices include specialized keyboards and interfaces designed to support individuals in message composition and speech synthesis. Composing a message with an AAC device can take more time than using verbal speech which poses social interaction challenges to AAC users. Augmented communicators (ACs) often fall behind in group conversations when topics move fast or only get asked yes/no questions instead of open-ended ones, limiting the AC's possible responses. While focusing on improving the rate at which AAC users can communicate and increasing ease to vocabulary access have been an important focus of the field of designing for AAC, this dissertation presents and argues for *conversational agency* as a new research and design frame to study AAC technology and AAC-based interactions among disabled and non-disabled individuals.

Conversational agency is shaped by social constraints such as when AAC users can speak, what they can say, and who they can address. The first part of this dissertation proposes metrics to measure how conversational agency is expressed within constraints specific to AAC-based interaction. Using agency as a lens for AAC research uncovered new design opportunities for AAC systems that center AAC users' personal conversational goals. The second part of this thesis explored technologies that can help balance participation in a group and signal to partners that an AC is still making use of their turn to address social constraints related to managing attention and conversational dynamics. The third part of this dissertation explores technologies that can retrieve content from a conversation and clarify misunderstandings to provide ACs with more control over the content of a conversation and the relevance and context of their messages. Through empirical studies, co-design, participatory design, and the development and deployment of different systems, this dissertation investigates how different design materials, such as expressive customized robots and Large Language Models, can support AAC users in easily exercising their conversational agency and help non-AAC users become better communication partners. Through six studies this dissertation contributes Agency in AAC as a new design framework to both further our understanding of AAC users agency and to generate novel, accessible, and agency-increasing communication tools.

Dedication

*To my parents who have never given up,
and to Ofelia, mi abuelita.
Gracias por tu amor constante
y por siempre creer en mi.*

Acknowledgements

I deeply appreciate the privilege of being able to pursue my doctoral studies. It has been a journey I owe first to God, for always providing a way when the roads seemed a little dark and for giving me strength. A journey also made possible to my family and friends and the mentors I met along the way who invested and believed in me so that I could do the same. My heart is filled with gratitude for all the wonderful individuals whose support has carried me through.

I want to thank my wonderful advisors Dr. Henny Admoni and Dr. Jeffrey Bigham. In addition to being incredible researchers, they are amazing human beings and it has been a wonderful opportunity to grow as an independent researcher under their mentorship. Thank you for supporting me with the freedom to explore this interdisciplinary space of AAC, agency and design. I also want to thank Dr. Jeff Higginbotham and Dr. Jodi Forlizzi, whose work has inspired me so much and for their continued support, guidance and mentorship.

I am also so thankful for Dr. Amy Pavel who has not only taught me how to be a better scientific communicator through giving me feedback on my posters, papers, and slides but has also been a fantastic collaborator who has modeled great mentorship skills. I have learned so much from you. I would also like to thank Patrick Carrington and Cindy Bennett who have been there for me as inspiring mentors committed to help me grow. I first met Patrick, Amy, and Cindy as postdocs and leaned on them for guidance. You really helped me thrive in Grad school and have always been there to help with career and research questions.

The work presented in this dissertation has been possible thanks to community members, collaborators, research participants and support networks both personal and academic. It really takes a village to develop research ideas and to recover from all the work. Thank you to Frederick Shic, Megan Smith, Laura Callinan, the Yale Child Study Center Lab, Shaun Kane, Merrie Ringel Morris, Mark Steidl, Tina Calabro, Sara Pyszka, Michal Luria, Michael Rivera, Lynn Kirabo, Lea Albaugh, Megan Hofmann, Samantha Reig, Julia Cambre, Michael Xieyang Liu, Tomohiro Nagashima, Vikram Kamath, Julian Ramos, Jill Lehman, Reuben Aronson, Franceska Xhakaj, Fannie Lu, Ben Newman, Abhijat Biswas, Ada Taylor, Pallavi Koppol, Michelle Zhao, Michael Lee, Maggie Collier, Tesca Fitzgerald, Cole Gleason, Mary Beth Kery, Prakhar Gupta, Sai Ganesh Swaminathan, Jared Santa Maria, Seunga Yu, Erica Cruz, Liz Carter, Jason Wu, Qian Yang, Kristin Williams, Udaya Lakshmi, Queenie Kravitz, Ryan Ries, Eric Davidson, Karen Harlan, John Davis, Emma Jiang, Jaeyeon Lee, Courtney Walker, the fantastic Atypical Interactions group, my Pittsburgh Familia, and many others. Last but not least, thank you Jesus Silva Elizalde for being my unwavering supporter, cheerleader, proof-reader y mejor co-equipero.

Published work from this dissertation

1. Stephanie Valencia, Amy Pavel, Jared Santa Maria, Seunga Yu, Jeffrey P. Bigham, and Henny Admoni. *Conversational agency in augmentative and alternative communication*. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pages 1–12, 2020.(CHI’20).

This work introduces the concept of Conversational Agency in AAC, the constraints impacting AAC users’ agency as well as metrics to measure agency during social interaction.

2. Stephanie Valencia, Michal Luria, Amy Pavel, Jeffrey P. Bigham, and Henny Admoni. *Co-designing socially assistive sidekicks for motion-based AAC*. In proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, HRI ’21, page 24–33, New York, NY, USA, 2021. Association for Computing Machinery.(HRI’21).

This work introduces the concept of Motion-based AAC, explored through a co-design workshop with augmented communicators and puppeteers.

3. Stephanie Valencia, Mark Steidl, Michael Rivera, Cynthia L. Bennett, Jeffrey P. Bigham, and Henny Admoni. *Aided Nonverbal Communication through Physical Expressive Objects*. In proceedings of the 2021 International ACM/SIGACCESS Conference on Computers and Accessibility.(ASSETS’21).

This work describes a case study design and evaluation of a co-designed “sidekick”, a motion-based AAC device that provides nonverbal communication for an augmented communicator.

4. Stephanie Valencia, Richard Cave, Kallarackal, Katie Seaver, Michael Terry, and Shaun K. Kane. *“The less I type, the better”:* *How AI Language Models can Enhance or Impede Communication for AAC Users*. In proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI 2023).

This work evaluates large language models as a tool to support AAC users in generating phrases and introduces *speech macros* as a method for AAC users to benefit from the generative capabilities of these models.

Portions of this dissertation were previously included in the scientific articles here listed. Permission to reproduce the text is included in the original publication’s copyright.

List of Abbreviations and Definitions

- **AAC:** Augmentative and Alternative Communication.
- **AC:** Augmented Communicator, AAC device user.
- **CCP:** Close Conversational Partner to an augmented communicator like a close friend or family member.
- **NLP:** Natural Language Processing.
- **Technical Agency:** Agency we express when we are able to participate in conversation. Taking a turn in a conversation is an example of a person displaying their technical agency.
- **Colloquial Agency:** Colloquial agency is achieved through technical agency when participating in a conversation leads to accomplishing an individual's goals.
- **SoA:** Sense of Agency. Self-reported subjective measure of agency.

Contents

1	Introduction	15
1.1	Agency as a Design Framework	16
1.1.1	Defining and Understanding Agency in AAC	16
1.1.2	Designing for agency	17
1.1.3	Designing with Agency	18
1.2	Overview and Contributions	19
2	Background and Related Work	21
2.1	Augmentative and Alternative Communication	21
2.1.1	Components of an AAC device	21
2.1.2	Social interaction challenges in AAC	22
2.1.3	Research Directions in AAC	23
2.2	Agency	23
2.2.1	Agency and Technology	24
2.2.2	Agency and AAC	24
2.2.3	Agency Challenges in AAC Interactions	24
2.2.4	Conversational Agency Design Framework	25
2.3	Technology as Design Material for AAC	26
2.3.1	Motion through Expressive Robotic Objects	27
2.3.2	Context Support through Language Technologies	27
2.3.3	Emerging Language Technologies for AAC	28
2.3.4	Augmenting Conversational Context Awareness in AAC	29
2.4	Design Frameworks in Accessibility Research	30
I	Agency as a Design Framework	32
3	Conversational Agency in AAC	33
3.1	Overview	33
3.2	Method	35
3.2.1	Tasks	35
3.2.2	Operationalizing Agency	36
3.2.3	Participants and data collection	37
3.2.4	Procedure	38
3.2.5	Study Limitations	39
3.3	Results: Constraints and Participation	39
3.3.1	Effect of Conversational Constraints on Participation	39
3.3.2	Effect of Relationship Constraints on Participation	42
3.3.3	Effect of Task Constraints on Participation	43
3.3.4	Effect of Device Constraints on Participation	44
3.4	Results: Participation and Technical Agency	44

3.4.1	Supporting Independent AC's Communication	44
3.4.2	Increasing Information Exchanged	45
3.4.3	Creating Missed Opportunities	45
3.5	Results: From Technical to Colloquial Agency	46
3.6	Discussion and Reflection	47
3.6.1	Designing for Conversational Constraints	47
3.6.2	Designing for Relationship, Device and Task Constraints	48
3.7	Summary	48

II Agency through Motion 49

4	Co-Designing forms of Motion-based AAC through Robotic Sidekicks	50
4.1	Overview	50
4.2	Method	51
4.2.1	Step 1: Online Survey	52
4.2.2	Step 2: Co-design Workshop	52
4.3	Workshop Findings	54
4.3.1	Attention, Precision, Timing: Parameters for Nonverbal Communication	54
4.3.2	Conversational Goals for Sidekicks	55
4.3.3	Sidekick Aesthetics & Use Cases	57
4.3.4	Sidekick Behaviors and Motions	58
4.4	Workshop Accessibility Reflections	59
4.5	Discussion & Future Work	60
4.5.1	Precision through Embodied Motion	60
4.5.2	Timing of Sidekick Behaviors	61
4.5.3	Modalities for Gaining Attention	61
4.5.4	Building Motion-Expressive Sidekicks	61
4.6	Summary	62
5	Aided Nonverbal Communication through Physical Expressive Objects	63
5.1	Overview	63
5.2	Design Team Participants	64
5.3	Design Process	64
5.3.1	Discovery: Opportunities for Expressive Objects in AAC	64
5.3.2	Definition: Goals and Sidekick Properties	65
5.3.3	Development: Ideating and Prototyping	66
5.4	Prototype Implementation	68
5.4.1	Initial programmed motions	69
5.5	Delivery: Testing and Iterating	69
5.6	Evaluation: Use and Impact	70
5.6.1	Sidekick Use and Performance	71
5.6.2	Close and New Communication Partners	72
5.6.3	Better in-person	73
5.7	Discussion	73
5.7.1	Benefits of aided nonverbal communication	73
5.7.2	Barriers to long-term use and AT development	74
5.7.3	Accessibility of the long-term co-design process	75
5.7.4	Limitations and Future work	75
5.8	Summary	76

III	Agency through Language	77
6	Discovering Opportunities for Language Technologies	78
6.1	Conversation Interaction Analysis	78
6.1.1	Analysis	79
6.1.2	Results	80
6.2	Interdisciplinary Perspectives on Language Technologies for AAC: A Participatory Design Workshop	80
6.2.1	Participants	81
6.2.2	Procedure	81
6.2.3	Data collection and analysis	82
6.3	Challenges and Opportunities for Aided Support	84
6.3.1	C1: Grammar and Pronunciation	84
6.3.2	C2: Misunderstandings and Repair	85
6.3.3	C3: Accessing New Vocabulary	87
6.3.4	C4: Partner Feedback	88
6.3.5	C5: Managing Turns and Timely Participation	89
6.4	Discussion: a Research Agenda for NLP in AAC	90
6.4.1	Using Conversational Context to Reduce Misunderstandings	90
6.4.2	Controlling Text Generation to Improve Expressivity	92
6.4.3	Limitations	93
6.4.4	Conclusion and Future Research	93
6.5	Summary	93
7	How Large Language Models can Support and Hinder Communication	95
7.1	Overview	95
7.2	Integrating LLMs into AAC with Speech Macros	96
7.2.1	Design Process	96
7.2.2	Speech Macros	97
7.2.3	Prototype User Interface	98
7.3	Evaluation Study	98
7.3.1	Participants	100
7.3.2	Procedure	100
7.3.3	Data Collection and Analysis	102
7.4	Findings	103
7.4.1	Benefits and Uses of Speech Macros	103
7.4.2	Learning Input Mappings	105
7.4.3	Evaluating Suggested Phrases	106
7.4.4	Concerns About Using AI Suggestions	109
7.5	Discussion	110
7.5.1	Can Speech Macros Improve AAC Use?	110
7.5.2	Design Challenges and Trade-offs	110
7.5.3	Conducting Human-Centered AI Research with AAC Users	112
7.5.4	Limitations and Future Work	112
7.6	Summary	113
8	Augmenting Context Grounding with the COMPA system	114
8.1	Overview	114
8.2	Design Goals	115
8.3	COMPA	116
8.3.1	Interfaces	116
8.4	User Study	117

8.4.1	Participants	119
8.4.2	Procedure	119
8.4.3	Analysis and Data Collection	120
8.4.4	Study Limitations	121
8.5	Findings	121
8.5.1	Using COMPA's Features in Conversation	121
8.5.2	Effects of COMPA's Features on Participation	123
8.6	Discussion and Future Work	126
8.7	Summary	127

IV Contributions, Future Work and Conclusion 128

9	Contributions, Future Work, and Conclusion	129
9.1	Contributions	129
9.1.1	Theoretical Contributions	129
9.1.2	Empirical Contributions	130
9.1.3	Technical and Design Contributions	130
9.1.4	Methodological Contributions	131
9.1.5	Agency Questions that Remain	132
9.2	Conclusions	133

List of Figures

1.1	Conversational constraints (<i>e.g.</i> , who can speak, who can be addressed, what is being said, what is expected of others), the task at hand, the relationship with other speakers, and the AAC device being used can all impact the expression of an augmented communicator’s agency in conversation. This dissertation, lays its foundation on how conversational agency* [54] is defined in the social sciences to propose Agency in AAC as a design framework that can describe how agency is expressed in AAC-based interactions, and how we can use agency as a lens to generate accessible communication tools.	18
3.1	Conversational constraints (<i>e.g.</i> , who can speak, who can be addressed), the task at hand, the relationship with other speakers, and the device being used can all impact the expression of an augmented communicator’s (AC) agency in conversation. . .	33
3.2	Conversational, task, relationship and device constraints can change the pathways for communication (arrows) between AC, CCPs and a third party (TP) and the effort, or cost, required to communicate along the pathways (arrow thickness). For example, time-sensitive conversational constraints (<i>e.g.</i> , replying fast while it is still relevant) can encourage communication along the lowest cost paths (lower width arrows) encouraging communication via their CCP rather than direct AC to TP communication. When this delegated technical agency enables the realization of goals for ACs we call it Colloquial Agency.	34
3.3	Study setup. The CCP (left), the AC (middle) and the experimenter(right) sit at a round table. A scene camera captures this image, a camera behind the AC captures activity on the AAC device, and a camera mounted on the AC’s wheelchair captures AC’s facial expressions.	35
3.4	The interview task mirrors open conversation except that the third party only directly addresses the AC (missing arrow between the third party and CCP) and only the AC and CCP share knowledge to answer the interview questions (small blue dots). In the first map task condition (C1), the AC and the third party can see the correct route (small blue dots) and the AC must communicate to the CCP what to draw on the map. In the second map condition, the AC and CCP share the correct route and must tell the third party what to draw on the map. In the craft task, the AC generates ideas for their craft which is largely constructed by the CCP (the third party does not participate).	36
3.5	Overall, ACs contributed a higher percentage of turns immediately after they were addressed by the third party (F). CCPs in pairs 1 and 2, contributed a higher percentage of the time in subsequent turns (S).	40
3.6	When CCPs took a turn during the interview, they most often gave explanations (E). CCPs also asked questions for confirmation by the AC (QC) and asked permission before providing explanations (P).	40

3.7	When ACs took a turn during the interview, they most often gave explanations (E). Other turns were spent confirming (COI) or denying (NOI) information given by the CCP, granting permission (GP) to CCPs to provide explanation, requesting help (RH) and requesting time (RT). ACs in P1 and P2 explained in less than 60% of turns, while ACs in P3 and P4 explained in more than 90% of turns.	41
3.8	Percentage of turns taken by the AC and CCP per task for each participant pair. .	42
3.9	Modes of communication used by AC per task	43
3.10	Example materials for the map and craft task. For the map task, we show an example guide map with route information (visible by AC) and follower map (visible by CCP or third party) that shows only the starting point, requiring the follower to draw the rest of the information (displayed in pen). For the craft task, we show the collages produced by each participant pair. Names that appear on collages are the intended receivers.	44
4.1	We explore the metaphor of an “assistive sidekick” to design robots that assist augmented communicators by playing a supporting role through nonverbal behavior.	51
4.2	Our two step process included formative work and a three-phase co-design workshop. Participants included augmented communicators (ACs), their close conversational partners, and professional puppeteers.	51
4.3	Prototypes by Tammy’s (left) and Matt’s teams (right). Tammy’s prototype featured a large set of eyes, customized to express sentiment. Matt’s prototype read “one moment please” to communicate a desire to contribute.	57
5.1	A design team including an augmented communicator, his close conversation partners and HCI researchers worked together on designing a physical expressive object to support AAC-based communication.	63
5.2	Expressive robotic objects ranging from character-like to object-like. (A) Tega, a robot emotive companion [164] image credits to Bruce Peterson, (B) Emoto, an emotive AI sidekick made of a robotic stand that acts as a body for your phone [19], and (C) Paper Signals, image credits to Google, animated objects that track and display data through embodied motion [18].	66
5.3	Low-fidelity prototype example motions	67
5.4	A) Early prototype 3D model; B) Front view and C) side view of fabricated sidekick attached to wheelchair mount.	69
5.5	Hand sketch produced by D explaining the new timer motion.	70
5.6	A) The timer motion moves 30 positions right to left and then 30 positions left to right, advancing 6 degrees per step; (B) The wave motion waves back and forth 6 times; and (C) the reset sequence centers the base and lowers the arm.	70
5.7	Parents (D, T) and brother (P) are included in the family circle. Friends include past aides (PA), AAC group friends (AC), music camp friends (WL) and neighbors (NH). Acquaintances include co-workers (CW), HCI collaborators (CL), and current aides (CA). Service providers include a job coach (JC), physical therapist (PT), Supports Coordinator (SC), and an assistant’s agency (BY). In the outermost circle a disability rights community (DA) has recently started to get to know Mark.	71

6.1	Challenge and opportunity areas uncovered in our participatory workshop along with names for solutions participants came up with. We highlight the challenge areas AC participants identified as bringing the most value to them. AC1 ranked <i>accessing new vocabulary</i> as his top choice and <i>misunderstandings and repair</i> as his second. AC2 was highly interested in tools to help him address past topics in conversation and ranked <i>misunderstanding and repair</i> as a priority area followed by <i>partner feedback</i> and <i>accessing new vocabulary</i>	83
7.1	Prototype of <i>speech macros</i> used in the study. Each screen includes the name of the current macro, the conversation partner’s question context, space for the user’s input, and a ‘variability’ slider to adjust the temperature of the model, and thus the diversity of the responses generated. Output from the LLM is presented on the right side of the screen.	99
7.2	Storyboard exemplifying a sample interaction with the Extend Reply Speech Macro.	101
7.3	Most participants found speech macros to be either very or extremely useful. Turning words into requests was considered very and extremely useful by most participants. Participants identified they could use the turn words into requests macro to help them daily to create requests for routine, self-care and accessibility related tasks.	106
8.1	COMPAs is a browser extension that enables Context Marking and Phrase Assistance to augment contextual grounding and awareness about what an AAC user may want to refer to in conversation. Using COMPAs, AAC users and communication partners can view a live transcript (A,D) of their conversation and notifications (C) about the context of a conversation an AAC user may want to reply to. COMPAs also explores how enabling AAC users to choose an intention when composing a message (<i>i.e.</i> , share an opinion or ask a question) can provide better starter phrase suggestions and more accurate communication partner notifications (B).	115
8.2	We created three versions of COMPAs. Each version added a new layer of context information and phrase assistance type. Testing these three different versions would allow us to evaluate the contribution of each grounding strategy to the conversation experience.	118
8.3	All study participants (N=8), four AAC users and four communication partners (CPs), rated all of COMPAs features after experiencing all versions of the tool. CPs rated pausing the transcript, marking the context, and general typing alerts as the most useful features to help them better communicate with their augmented partners. All AAC users rated the general typing alerts and the intent-specific typing alerts as highly useful to them.	122
8.4	AAC users rated each in version in terms of how likely they would use it again if it were available to them. Most AAC users rated COMPAs’s second version as the version they would most likely use while CPs selected the full version of COMPAs as their preferred version.	123
8.5	AAC users stated they were able to always (5, rating value) or often (4, rating value) participate as desired when using the second version of COMPAs. AAC users stated they were either often (4, rating value) or sometimes (3, rating value) able to participate as desired with COMPAs 1 and COMPAs’s full version. AAC users felt they participated much more when using the full version of COMPAs (COMPAs 3) compared to the other two versions.	125

8.6 AAC users initiated conversation more often when using the full version of COMPA (18 total initiations by AAC users). COMPA’s full version enabled AAC users to retrieve intent-specific phrase suggestions and share intent-specific typing notifications with their partners. 125

List of Tables

1.1	The four conversational constraints that impact agency according to Gibson, what governs the constraint, and factors that contribute to the constraint application. . .	17
3.1	We summarize the four conversational constraints that impact agency according to Gibson, what governs the constraint, and factors that contribute to the constraint application.	34
3.2	Our eight study participants, four augmented communicators (ACs) and their corresponding close conversation partners (CCPs). Using our transcripts, we calculated the throughput or total words spoken over minutes typing (WPM) for each AC.	38
3.3	Collage elements in the final craft that were proposed by the AC, proposed by the CCP and agreed to by the AC, or proposed by the CCP and included in the artifact without explicit agreement from the AC.	46
4.1	Goals along with their corresponding numbers and the stage at which they were introduced: original goal from formative work (O), goal mentioned during the <i>telling</i> phase (T), goal mentioned in <i>making</i> (M). Goals 1, 5, 8, and 10 were selected to be further explored in the <i>enacting</i> phase.	56
4.2	Puppeteers generated 10 behaviors for each selected goal using sidekick prototypes of their choice. We highlight three example behaviors per goal.	58
5.1	List of meetings reported with diary entries over a period of three weeks.	72
6.1	Participants included expert augmented communicators (AC1, AC2) with more than 10 years of experience using speech-generating AAC devices; Communication Scientists some who were also certified speech-language pathologists (SLPs) (COM1, COM2, COM3), and technologists specializing in natural language processing and generation (NLP1, NLP2, NLP3) and Human-Computer Interaction and accessibility (HCI1, HCI2, HCI3).	82
7.1	Our 12 study participants (3 females, 2 non-binary (NB)) used a diverse set of AAC devices and techniques and had different levels of speech use.	100
7.2	Extend Reply Speech Macro inputs given by our study participants, the suggestions returned by the speech macro, their preferred option (shown in bold), and their rating of how appropriate that suggestion would be for the question asked and their personal communication style.	104
7.3	Participants suggested different items to be turned into help requests. Items spanned three main categories: medical, accessibility related, and daily tasks.	108

8.1 Our eight study participants included four AAC user study participants (2 females) and four corresponding communication partners. AAC users have 10 or more years of experience using AAC and use a diverse set of AAC devices and access methods. 117

8.2 Participants tried three COMPA versions that had different levels of context information for each of COMPA’s features. Participants tried all versions in a randomized counter-balanced order. 120

8.3 Average turns taken and words per minute spoken by turn. The average number of turns taken by AAC users and CPs was comparable when participants used the full version of COMPA. 124

Chapter 1

Introduction

Communication is essential to agency. Agency in conversation, is defined as the ability to advance our individual goals by communicating them [54]. Nonetheless agency is not a fixed property; it manifests within structure [41] and social constraints [54, 42] that challenge its expression. Augmentative and alternative communication (AAC) devices can support conversation for people with speech and language disabilities [14] but communicating with one of these devices poses unique challenges to a users' agency. Many augmented communicators (ACs) who use speech generating communication devices encounter social constraints and barriers that reduce their agency in conversation including how much they can say, how they can say it, and when they can say it. Conversational partners [15] and social constraints embedded in the traditional ways we communicate [54] can impact the quality of an interaction by either facilitating or hindering the expression of an AC's agency in conversation. While social theory has described how agency is shaped and expressed in conversations between oral speakers, this dissertation focuses on studying agency in AAC. Through this work, agency in AAC emerges as a new design framework that can be used to study agency in AAC and generate new accessible and agency-increasing AAC tools.

Using an AAC device to speak often takes time, with rates ranging from 3 to 20 words per minute [65, 77] which creates a time asymmetry between an AC and their conversation partners who use their mouth and body to speak. For example, when a communication partner knows that it takes an AC more time to respond, the communication partner may ask yes/no questions instead of open-ended ones, limiting the AC's possible responses [66]. Conversation partners can also dominate the conversation, or fail to respond to communication attempts by ACs [62]. Such conversational partner impact on ACs' participation can be uncomfortable for ACs [46] and can reduce perceived autonomy and agency [62]. Prior work has identified challenges in interactions between augmented and non-augmented conversants but has not yet determined how different types of conversational partners impact an AC's conversational agency, what other factors impact ACs agency, and how we can design technology that support's ACs' agency expression. Futhermore, little is known about how specific AAC interface designs, existing and future designs, may shape agency in conversations among non-AAC and AAC users.

As a first step towards making communication possible and efficient, designers and researchers of AAC have focused on improving device throughput, for example via text suggestions using language technology [167], and vocabulary sorting [5, 78]. Human-computer interaction (HCI) researchers have also worked on improving expressivity by adding partner-facing displays [146], and customizable voices to speech generating AAC devices [45, 104, 121, 118]. Research from communication science has focused on increasing communicative competence of ACs [90, 92] and teaching conversation management strategies such as learning how to request partners to wait for a response and using pre-composed messages to save time when sharing a thought in conversation. Research in Conversational Analysis, the microanalytic study of social interaction, has contributed rich descriptions on how AAC and non-AAC users communicate with each other [66] and take turns [131]. This dissertation extends this foundational research to uncover how technology and

design choices can enforce or loosen social constraints that impact an AC's agency.

HCI researchers have recognized the role communication partners play in hindering or facilitating an AC's communication. Ibrahim et al. studied how a conversation partner's assumptions about an AC's competence could negatively impact children's agency when using their AAC device to communicate at school [70]. Fiannaca et al. recognized that CCPs could facilitate an AC's communication through informed suggestions and designed a companion app to enable CCPs to suggest words to help their augmented partner as they composed a message [46]. While there has been a development on new technology for AAC there is still no model of how these new developments can increase or decrease agency for ACs. There is also no description on how constraints specific to AAC such as partner's level of experience, their relationship to the AAC user, AAC devices, and tasks, can impact how ACs can express their agency. In this dissertation, I focus on closely understanding social interaction among AAC and non-AAC users to translate the observed behaviors that impact how ACs can express themselves into an informative design framework for agency in AAC.

Building from this formative understanding of agency with AAC, this dissertation goes on to investigate: What factors impact agency in AAC? How can we measure agency in AAC? What technology materials and interactions can increase the opportunities for ACs to express their agency? Specifically, how can we leverage specific modalities and technology platforms to create agency-increasing tools for AAC? What system features support or hinder user agency? What are appropriate research and inclusive design methods to explore these possible new technologies with AAC users and other relevant stakeholders?

1.1 Agency as a Design Framework

This dissertation focuses on first establishing conversational agency in AAC as a descriptive and generative design framework through the definition of participation and communication-based metrics that impact how, when, what, and how much ACs can say in conversation. Using this design framework, this thesis then continues exploring opportunities to loosen the social constraints identified and proposes new designs through collaborative and participatory methods that include ACs in the design process.

1.1.1 Defining and Understanding Agency in AAC

To define agency based on ACs' experiences, this dissertation build on Gibson's work on conversational agency from sociology [54]. Conversational agency can be measured as (1) participation in conversation, also known as *technical agency* and (2) achieving a particular goal because of that participation, known as *colloquial agency*. Gibson's theory states that there are different conversational constraints that act as social rules that can impact the expression of agency. These conversational constraints include who can be addressed (*participation shift constraint*), what type of response we expect from others (*the ritual constraint*), the topic of conversation (*relevance constraint*), and when is the right time to start speaking (*one speaker constraint*). These conversational constraints along with the factors that govern them are summarized in Table 1.1.

While these constraints are described for conversations among individuals who do not use AAC, in **Part I** of this dissertation, we describe a first study designed to understand how conversational agency is expressed in the AAC context. The outcome of this first study is a framework for conversational agency in AAC that considers how conversational constraints and AAC-specific constraints such as the relationship with their conversation partners, the AAC device of use, and the conversational task they are carrying out (e.g., a formal meeting with goal, an open-ended chat) can impact ACs' agency (Figure 1.1).

This first study included ACs and their close conversational partners (CCPs) described in detail in Chapter 3. CCPs such as parents and paid aides, who have caregiving roles, often support

ACs by facilitating conversations with others [15]. I considered scenarios in which there is an unfamiliar third party (TP) to study how CCPs facilitate the conversation between the AC and the TP, and how this impacts the ACs’ agency in the context of specific conversational tasks.

Table 1.1: The four conversational constraints that impact agency according to Gibson, what governs the constraint, and factors that contribute to the constraint application.

Constraints	Governs	Contributing factors
One speaker	Current speaker	Time to start speaking
Participation shift	Next speaker	Addressing next speaker (<i>e.g.</i> by name)
Relevance	Topic of contribution	Time spent speaking without pause
Ritual	Type of contribution	Previous speech act (<i>e.g.</i> question)

The first study findings showed that ACs participated more frequently when directly addressed in an interview task than in subsequent turns when CCPs often stepped-in to comment. That means that in order to provide ACs with more control over the participation-shift constraint future work could consider how to ensure speakers are addressed or can signal wanting to participate equally through technology or conversational methods.

1.1.2 Designing for agency

Extending the findings from this first study, this dissertation then employs the conversational agency in AAC framework (Figure 1.1) to understand how we can design and deploy agency-increasing technology for ACs that focuses on loosening specific conversational constraints. This dissertation explored two different technology materials¹ composed of different modalities and technology platforms to study how they can be used to increase agency in AAC.

In **Part II**, *Agency through Motion*, I explore how physical and expressive objects and robots can support users in managing conversation dynamics, a salient area of opportunity in the Agency Framework 1.1. Chapters 4 and 5 explore physical design materials that can support bringing attention to an AC, manage turn-taking, and holding the conversational floor with the goal of addressing the participation-shift and one speaker constraints. We learned that motion could be used to bring attention, convey precise messages and support timely participation. Attention, precision, and timing all contributed to increasing AC’s agency in conversation with unfamiliar and close conversation partners [156].

In **Part III**, *Agency through Language*, this dissertation then focuses on exploring language technologies as a design material that could potentially support AAC users in managing a conversation’s topic and the way words are phrased which could help address the relevance and ritual constraint. This part of the dissertation explored different ways to support establishing mutual understanding among conversants despite differences in response time among AAC and non-AAC users. Natural Language Processing (NLP) techniques and machine learning can support ACs in addressing desired topics in conversation even if they have already passed. Language technologies can keep track of a conversation, record it and process it in a way that can be accessed by ACs to address topics that they desire to revisit. NLP can also be used to detect parts of speech that can be useful in helping an AC’s respond to specific requests made by a partner or other speech acts. Using this material’s functionalities we set to identify opportunities to support AC’s agency

¹Throughout the thesis, I will purposefully switch between the term “technology material” and “design material” to denote the output modalities and technology platforms I choose to work with based on their potential to support agency in AAC. This definition comes from the design approach to reflect on the materials a designer uses when creating something new [135, 165].

through a participatory and interdisciplinary workshop (Chapter 6) and explore the opportunities in depth with two additional studies, one focusing on large language models (Chapter 7) and the last study on the design and evaluation of a tool to augment contextual grounding for AAC users (Chapter 8).

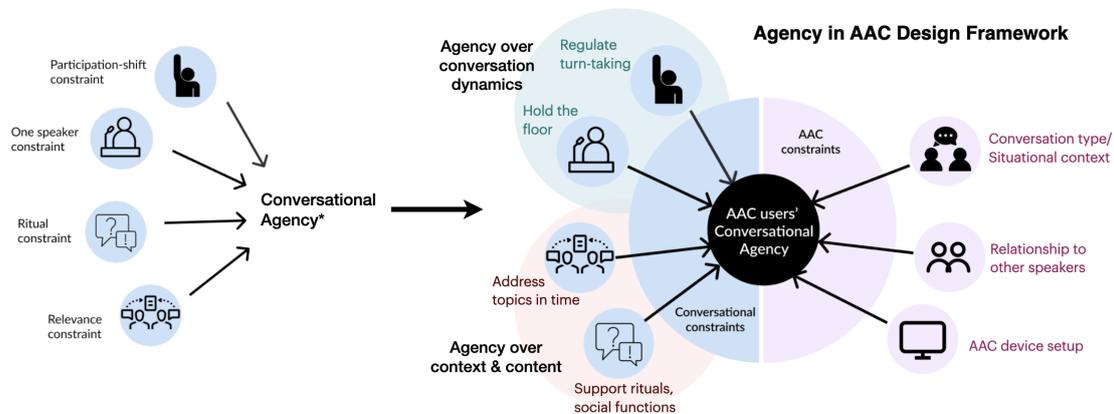


Figure 1.1: Conversational constraints (*e.g.*, who can speak, who can be addressed, what is being said, what is expected of others), the task at hand, the relationship with other speakers, and the AAC device being used can all impact the expression of an augmented communicator’s agency in conversation. This dissertation, lays its foundation on how conversational agency* [54] is defined in the social sciences to propose Agency in AAC as a design framework that can describe how agency is expressed in AAC-based interactions, and how we can use agency as a lens to generate accessible communication tools.

1.1.3 Designing with Agency

This dissertation not only focuses on how we as technology designers and researchers can increase user agency through artifacts but also on how we can increase the agency of AAC users in the design process of new technology. Throughout **Part II** and **Part III** I use design research methods such as co-design [128] and research through design [172] to explore new technology materials for AAC. I have adapted participatory user research methods to be accessible to ACs by making practical modifications such as facilitating asynchronous and remote ideation, easy referencing of abstract concept ideas on the go, and getting feedback from users at different stages of design.

This dissertation also includes two participatory design workshops that were carefully designed to scaffold discussions among diverse stakeholders (Chapter 4 and 6). Participatory workshops have shown to be productive in developing a congruent view of the problem space in areas of accessibility that require an interdisciplinary approach [20]. Participatory design seeks to bring multiple experts, including end users to design together and elicit new research directions for AAC through selected co-design activities [128, 129]. The participation of distinct stakeholders in the design process is valuable in enabling interdisciplinary design thinking for problem-solving [37, 121]. A participatory approach can help us avoid making assumptions about what technology solutions would bring value to users which leads to many products that are not usable and quickly abandoned [120, 86, 67].

In this way, a core component of this dissertation focuses on “thinking through making” to complement our understanding of agency as an observed action or a subjective experience. Through understanding how agency is expressed by examining social interaction we can then inform the design of agency-increasing AAC tools and through the making these tools, in close collaboration with ACs and other relevant stakeholders, we can learn more about agency, accessibility

and inclusion. Underlying these design investigations with current and emerging technologies included in this body of work, is the thesis:

We increase conversational agency and participation in conversation for augmented communicators when we co-design tools that address conversational constraints.

1.2 Overview and Contributions

This dissertation is organized in three parts that progress from the starting question “How does agency manifest in AAC?” to the concrete experiences of designing AAC applications that increase conversational agency for augmented communicators.

Chapter 2 provides the background including an overview of AAC, how agency has been defined and studied in the field of HCI and the social sciences, definitions of conversational agency and technology materials that have been used to augment AAC systems.

In **Part I**, chapter 3 describes how ACs express their conversational agency through an observational study to understand how conversational agency manifests in the face of constraints specific to AAC. This chapter answers the following research questions: *How is conversational agency expressed in AAC? How can we quantify agency in AAC? What factors impact an AC’s agency?* The chapter concludes by outlining a design space of conversational agency in AAC and suggests possible ways to address the constraints presented in the framework.

In the process of learning more about conversational agency in AAC and what technology materials contribute to it, Part II and Part III seek to answer the following main research questions: How can different technology materials address conversational constraints? How do the designed tools impact AC’s perceived and observed agency? **Part II** presents two studies that explore motion through expressive robotic objects as a design material to support AAC users in expressing their agency. Chapter 4 describes a co-design process to explore the opportunities for expressive robotic objects to support AAC users’ agency. Chapter 5 presents a case study detailing the co-design process and evaluation of a robotic expressive object with an augmented communicator to understand How socially assistive robots can address conversational constraints.

Part III focuses on addressing the relevance and ritual constraints using language technologies, such as large language models (LLMs), text-to-speech and interactive add-on digital tools. Chapter 6 explores different possible future technology concepts with ACs, NLP researchers and communication scientist. Chapter 7 supported ACs’ experimentation with LLMs by introducing the concept of *speech macros*: LLM prompts that transform abbreviated user input into full sentences, with a focus on achieving conversational goals such as requesting help with something or answering a biographical question. Chapter 8 presents COMPA a system that integrates different support strategies for AAC users and communication partners to augment their mutual understanding during real-time group meetings.

Finally, **Part IV**, containing Chapter 9, summarizes the contributions of this dissertation and reflects on this dissertation’s goal to describe how social and material constraints impact individuals’ ability to express their agency when using AAC.

In total this dissertation comprises six studies that altogether make the following contributions:

- Conversational agency as a theoretical foundation for investigating how social and technology constraints impact how augmented communicators participate in conversation and as a design framework to explore how to support users’ conversational agency (chapter 3).
- Accessible participatory methods to envision new technologies with Augmented communicators (Chapter 4, Chapter 5, Chapter 6).
- Design knowledge on how specific technology materials (modalities and platforms) can support conversational agency in AAC. (Chapters 4, 5, 6, 7, 8)

- Three systems and design concepts to support AC's agency. (1) Aided motion-based AAC through Sidekicks, expressive socially assistive robots that augment nonverbal communication (Chapter 5). (2) Speech Macros, which leverages user brief user input and pre-set instructions to a Large Language Models to aid in communication (Chapter 7). (3) COMPA, a Context Marking and Phrase Assistance AAC add-on digital tool to augment contextual grounding during online real-time group meetings (Chapter 8).

Chapter 2

Background and Related Work

This chapter draws on three bodies of research: The first focuses on augmentative and alternative communication (AAC). The second presents the conceptualization of agency in technology and disability diving deeper into agency challenges in AAC and interaction and how we can use Conversational Agency theory to create a design framework for AAC. The third overviews different AAC systems that have been developed to enhance AC's interactions and support AC's communication. In this third section I particularly outline what output modalities and technology platforms have been used to create augmentations to AAC systems.

2.1 Augmentative and Alternative Communication

Augmentative and Alternative Communication (AAC) encompasses all types of communication we use to express ourselves in addition to speech. AAC includes facial expressions, sign language, gestures, written language, and even Morse code [3]. Aided forms of AAC use technology to support alternative communication modes, such as picture or letter boards that people can point to or speech generating devices people can use to compose messages [14, 50]. Approximately 5 million Americans and 97 million persons worldwide may benefit from AAC [12]. In my dissertation work I specifically work with adult augmented communicators (ACs) who use speech generating devices and have motor disabilities due to cerebral palsy. As such, in this section, we provide an overview of speech-generating AAC systems.

2.1.1 Components of an AAC device

Some users who use AAC speech generating devices and have a mobility disability use stand-alone AAC devices that are mounted to their wheelchairs. Different AAC device properties can vary by user and can have an impact on their word per minute rate. One of these properties is the access method which can be either direct selection (user can directly choose words or letters) or indirect selection (user uses a scanning system that aids in selection of desired word or letter). Speech-generating AAC devices include different layers of technologies customized to fit a user's access needs including: accessible input methods, display and layout designs, encoding and prediction techniques, and speech and visual output features [13].

Accessible input methods

Users with motor disabilities commonly access AAC systems via indirect selection methods via a switch or an eye tracker. For switch-based input methods, the AAC device scan through each option and once the scanner is on the desired target a user presses a switch for selection. Users who are able to control the position of their head may opt for gaze-based access methods that

employ eye trackers for target selection. Gaze-based systems as well as switch scanning require considerable skill achieved through practice and training. Direct selection access methods are also a possible access method in some cases and refer to directly selecting a target with any body part, stylus, or joystick. Scanning access methods tend to have lower entry rates than other methods but in general High-tech AAC users with motor, disabilities may communicate at 3 - 25 words per minute (WPM) [157]. Although access methods are tightly linked to hardware, in the scope of this paper we will reflect on how language technologies can assist users depending on their access mode and how their access mode may inform design considerations for the implementation of different approaches.

Displays and layouts

AAC devices can have static or dynamic displays usually organized in grids that display different words or letters. These displays usually have different pages organized by activity, in alphabetical order, or idiosyncratically. Users have to navigate to each page to perform to retrieve specific information. Some newer commercial devices are starting to leverage language models to dynamically update the possible word options based on what a user is typing and continue to learn from each user's style (<https://spokenaac.com/>). Typical components in an AAC display include some sort of keyword and word or phrase buttons that have been either pre-stored or pre-configured. Different approaches to displaying and retrieving vocabulary require a lot of mastery and practice on behalf of users. AAC scholars have highlighted how current AAC systems place all effort on the human actor while the technology only stores linguistic items to be later retrieved [161]. In this work, we reflect on what portions of the communication process could be off-loaded to language technologies and reflect on the possible trade-offs this may incur.

Output modalities

Synthesized or digitized speech is the main output modality for high-tech AAC systems. Additionally, text-based partner-facing displays have been explored in research [146] and included in commercial systems (Tobii Dynavox, Partner Window) to help communication partners know when and what the AAC user is typing. Researchers have also explored motion as an additional modality to speech-generating AAC systems with the purpose of alerting other partners about the AC's speaking turn [156, 159]. In this work, we discuss how designing with new language technologies may require thinking about additional output modalities that may support both AAC users and communication partners in mutual understanding and extended social interaction.

2.1.2 Social interaction challenges in AAC

When communicating using an AAC device, people with speech disabilities have to first compose a message by typing out individual characters or selecting predicted words or icons before they share their thoughts. Speaking using an AAC device is slow compared to typical speech, with rates ranging from 3-10 words per minute [65, 77]. Having asymmetrical speaking rates [136] brings unique constraints to both the augmented communicator and their conversation partners. For example, when a communication partner knows that it takes an AC more time to respond, the communication partner may ask yes/no questions instead of open-ended ones, limiting the AC's possible responses [66]. Non-augmented partners can also dominate the conversation or fail to respond to ACs' communication attempts [46, 62, 157].

AAC users often have to decide how much they want to say and consider both the physical effort and the amount of time they will require before they decide if they want to compose their message [77]. It is then often the case that AAC users participate much less than their non-AAC user counterparts in conversation, are left behind in group conversation, or struggle to demonstrate

the relevance of their comment that is shared some minutes delayed after their topic they are addressing has passed [125, 136, 158].

Prior work has not yet determined a way to quantify and define how different conversation partner's may impact augmented communicator's participation in conversation and how they express their agency despite AAC device constraints and social constraints related to how different types of conversational partners impact an AAC users' agency. There is still no model of how constraints specific to AAC such impact the expression of agency in AAC in-person interactions. In the following sub-sections I will review how agency has been defined in the context of conversation and AAC and introduce a theoretical framework I build on to develop a Conversational Agency framework for AAC. This current thesis also contributes empirical knowledge on what social and material constraints impact an AAC users' agency in social interaction.

2.1.3 Research Directions in AAC

Current AAC systems demand users to be capable of completing many complex tasks. To exemplify, vocabulary retrieval in AAC is still largely done manually with the users needing to physically find the right words while being aware of the ongoing conversation at the same time. [161, 93]. Even though AAC systems benefit augmented communicators greatly, their design could still be improved to further support extended social interactions. There is a growing perspective shift in the AAC community from demanding AAC users to have a certain strategic, linguistic, and operational competence to operate their device to a perspective where AAC systems should take on more of these tasks [93]. The community has become more aware as to how AAC systems may impact important social interaction metrics beyond communication rate such as relational maintenance [35], conversational agency [157], and staying in time with a partner [66]. In this work we reflect on how language technologies could support users in achieving conversational goals related to social interaction and other valuable areas that are still unmet need and have been identified by users themselves.

Information from a conversation partner's speech has been used to generate contextually relevant topics [36], context-aware sentences and keywords [138], and collaborative input through a companion app [46]. Other approaches have leveraged a partner's speech to extract noun phrases and suggest them on an AAC device to be used by the AAC user [166]. Information about the characteristics of an AAC user has also been used to generate context-aware sentences by leveraging language models such as GPT-2 and context information included in AAC user persona tags and dialogue history [138]. Prompting large language models with subtasks have also been suggested to be useful in supporting assisted text entry. [169].

In 2012 a survey on NLP applications for AAC forecasted that future contributions to AAC from NLP would include spell-checking, context-based thesaurus, and genre-based word prediction [64]. Now ten years later we provide a participatory and interdisciplinary landscape of challenges and opportunities in AAC for current state-of-the-art language technologies. While systematic reviews about the current research in AAC [34, 64] and accessibility [99] provide a great overview of the field and current research directions in AAC, with our participatory approach we hope to open the conversation about future directions for AAC research by including ACs and communication scientists who do not often participate in computing conference but whose expertise can help paint a more complete picture of the problem space.

2.2 Agency

This section sets the theoretical background on agency and how it relates to AAC. I focus on how individuals with disabilities who use AAC systems express their agency when using technology-mediated communication and highlight some gaps in the literature that I plan to address with my dissertation work such as: How do augmented communicator make use of their agency in

conversation? How do AAC systems and people support or hinder this process? and how can we design agency-increasing AAC technology? What design materials (modalities and technology platforms) can be used to increase agency in AAC?

2.2.1 Agency and Technology

The field of Human-computer Interaction (HCI) has generally defined agency as a sense of control over the external environment and one's body [94]. Shneiderman's Eight Golden Rules of Interface Design [141] encourages interfaces that "support an internal locus of control," inspiring work evaluating perceived agency for interactions. In particular, such work considers implicit "sense of agency" measures (*e.g.*, perceived time of action and outcome) to study the impact of interface feedback [33]. In my dissertation, I examine matters of agency beyond the sense of control during an individual experience with an interface, and into matters of agency as a social experience, impacted in relationship with context and others [42, 43].

2.2.2 Agency and AAC

Goodwin, who studied AAC use with people with aphasia, notes that agency for ACs occurs within a social context where the meaning of communication must be formed cooperatively [57, 58]. Goodwin points out one issue that has emerged in the study of agency in disability, the difference between individual and "social" agency. Goodwin states that social agency is the one lodged within a social group, organized in an ongoing process of social cooperation [58]. In some situations, especially those in which a speaker relies on others to cooperate with them to be able to communicate, like in the case of a person who can only communicate with a limited vocabulary due to a speech disorder, this social agency is specifically salient. Goodwin describes the case of Chil, an individual who at age 65 suffered a brain aneurism making him unable to utter more than a few syllables and left him with little motor skills. Chil performed actions to communicate with others and express his desires that included pointing at meaningful references in his environment and making prosodic changes to his "no" and "yes" to guide his listeners to the right meaning as they asked different questions. Goodwin makes the point that human action has an intrinsically distributed organization in that it is constructed through the cooperative semiosis (signs and activities that carry meaning) of multiple actors [57].

This dissertation builds upon this work by studying how close conversation partners such as caregivers, friends or partners who know an AC well and unfamiliar conversation partners form meaning from AC's communication. In particular, I investigate how such support increases or hinders ACs' agency across different conversational tasks. Within this definition of social agency, this dissertation also focuses on co-designing technology that supports social conversational goals defined by AAC users themselves.

2.2.3 Agency Challenges in AAC Interactions

Prior observational [57, 58, 70], survey [46], and interview [4, 78, 77] studies have investigated AAC conversations among augmented communicators and close conversational partners (CCPs) identifying challenges to CCP-assisted conversation.

Ibrahim *et al.* observed children using their AAC devices to communicate with adults and their peers. They found that adults erroneously corrected the children's speech when they assumed children's competence using the device was low [70]. Adults also recounted a child's prior communicative act by accessing and reading out a previously constructed utterance with the AAC device without the child's consent. Fiannaca *et al.* surveyed ACs and their caregivers about caregiver-assisted communication, finding that the caregiver assistance can be uncomfortable when conversing with groups, and frustrating when the caregiver guesses what an AC wants

to say incorrectly [46]. Adults who adopted AAC devices later in life (due to Amyotrophic Lateral Sclerosis, ALS) reported becoming passive speakers who consider the cost of communicating when deciding to participate [77]. This suggests that agency may be also limited by the requirements needed to operate a device, such as energy and effort.

Communication scientists have also demonstrated how due to long waiting times associated with device operation in face-to-face conversation, ACs' conversation partners are more likely to ask yes-no questions instead of open-ended questions, dominate the conversation and fail to respond to communication attempts made by ACs [46, 65, 66]. This cited body of work shows that operating an AAC device brings in time pressure constraints that can have an impact on how much they can say and how often they are fully heard and that the AAC user's communication partner, plays an important role in influencing the quality of the conversational experience.

To illustrate the latter, Clark and Brennan [30] introduce the concept of the Least Collaborative Effort. In their work on grounding (i.e. the process of establishing what is being said and understood) they explain that people will not put much effort in trying to be understood by the other person if they do not have to. In the case of ACs, the Least Collaborative Effort can really have an impact on how much they are being heard or truly understood by their conversation partner. The literature shows several examples of the least collaborative effort in action. ACs' conversation partners exhibiting frustration with the time it takes to establish grounding with the AAC user and how in response to this frustration they will come up with interaction rules to speed up the conversation such as "Just nod, do not type on your device" –limiting the AC's actions [62, 63]. The conversation partner in this scenario may be introducing more error and barriers to the common effort of trying to establish common ground.

2.2.4 Conversational Agency Design Framework

As such, prior work has pointed out different factors that may impact an augmented communicator's participation in a conversation. Agency is modulated by a social context and impacted by others. This is especially salient in face-to-face interactions where speakers take turns and attend to each other's contributions in conversation. Agency is expressed through actions that further an individual's objective in the face of outside structure or constraints [30, 41, 55, 127]. Gibson, in his work on Conversational Agency, distinguishes two types of agency. *Technical agency*, is exercised when individuals contribute to the conversation (Figure 3.2). As individuals express technical agency by participating they can then advance their objectives in conversation, which he identifies as *colloquial agency* [54]. According to Gibson, an individual's technical agency can be promoted or reduced by four core conversational constraints. I define these constraints and describe how they matter for AAC conversations.

The ***one-speaker constraint*** specifies that only one speaker can speak at a time. When one speaker cues that they are almost done, another actor can preemptively start speaking to secure the floor. The one-speaker constraint implies that turn changes are a valuable resource for securing the opportunity for technical agency. When using an AAC device, ACs are at a disadvantage. Non-augmented speakers can secure a turn faster by speaking directly, while ACs have to operate their device to do this. Prior work has highlighted that ACs have challenges with social cuing or indications of turn-taking, as their speakers fail to recognize when ACs want to contribute to conversation [77].

The ***relevance constraint*** states that a new contribution to a conversation will be interpreted as a response to the most recent prior contribution. Thus, the relevance of a particular response decreases as the conversation proceeds. Together with the one-speaker constraint, the relevance constraint suggests that a very short window of time is available for an AC to make a relevant contribution to the conversation.

The ***participation shift constraint*** states that the target of a prior remark has the greatest odds of speaking next, but any participant may speak next. The prior target's advantage builds

as speaking provides other participants more opportunities for engagement. A known problem in AAC conversation is that ACs are often not addressed directly by third parties, instead third parties talk to them via their CCP. CCPs are recommended to help third parties engage directly with the AC by explaining that they can communicate directly [16].

The *ritual constraint* suggests conversation participants seek to maintain a positive social standing by showing considerateness to the speaker, assisting them with repair, and eliciting sympathy to them. Either through experience, exposure or training, CCPs may have specific ways in which they support their ACs' communication [16]. We expect to see these rituals, such as waiting for a response or assisting in a particular way, to play a supportive role for ACs to exercise their agency in conversation.

In this thesis, I use conversational agency as a frame to study AAC user's experiences in conversation. In addition to conversational constraints Gibson presents I consider additional constraints specific to AAC-based interactions such as the conversational task at hand (e.g., collaborative, open-ended), the relationship with other speakers (e.g., time knowing each other, comfort level), and the device being used (e.g., throughput and speech output types) can all impact the expression of an augmented communicator's (AC) agency in conversation. This framework is explained in detail in Chapter 3.

To understand participation in conversations and quantify metrics that show successful interaction, communication scientists studying AAC have used methods such as conversational micro-analysis, examining interactions at the utterance level [31, 32, 63, 91, 121]. I use similar micro-analysis methods to quantify participation by identifying specific occurring communicative functions (i.e., requesting more time, requesting help). More details can be found in Chapter 3. My dissertation adds to this body of work by analyzing participation across a set of different distinct conversational tasks to specifically investigate the relationship between CCP participation and ACs' expression of agency.

2.3 Technology as Design Material for AAC

The practice of HCI design involves engaging in reflective conversations with technology as a design material to create something new [135, 165]. Designers view technology as a design material to understand how the technology opens up and constraint design possibilities in a problem space and then use the material to envision things that have never existed before [135, 170]. Taking this technology material lens can inform design practice and can facilitate sharing design knowledge across disciplines [173, 170].

In this section, I first examine the technology materials that have been used to augment AAC systems and enhance conversations between augmented communicators and their conversation partners. Second, I provide some background on the specific technology materials I have identified have potential to be used for supporting conversational agency in AAC, these are (1) motion through expressive robotic objects and (2) natural language processing and machine learning techniques.

Designers and researchers have explored different design materials to augment AAC systems such as intelligent materials that can predict vocabulary from a conversation partner's speech [167] or from a video scene [5] or photograph [78]. HCI researches have also focused on designing for the AC's conversation partner to improve their awareness through partner facing digital screens known as awareness displays [146] and partner-facing companion apps to aid AC's in message composition [46]. In order to augment AAC systems with more expressivity, researchers have created customizable keyboards that can set specific voices to ACs' messages [45, 104]. Designers have also explored 17 different ways to say yes using buttons and different custom audio clips [118].

I have identified opportunities to explore two additional design materials that have the potential of addressing agency challenges in AAC tied to the four conversational constraints presented

in the conversational agency framework (figure 1.1). The first material has the potential of supporting ACs' in managing the participation-shift constraint and the one-speaker constraint through managing their partner's attention with embodied motion through an expressive object. The second material has the potential of supporting ACs' in addressing the relevance constraint and ritual constraint which relate to the content and style of a conversation. This second material involves machine learning and language technologies that have become a popular new material that communication scientists wish to explore for AAC [137].

2.3.1 Motion through Expressive Robotic Objects

Commercial speech generating AAC systems are currently only customizable at the word selection and speech production levels, and they do not yet support augmentations that can increase nonverbal communication. Nonverbal communication is key in helping regulate turn-taking, convey personality, and execute actions that increase social agency [81]. Gaze, body orientation, and pointing gestures are all examples of nonverbal communication that can support directing attention to different speakers in a group conversation [144, 81] and are especially powerful in face-to-face interactions. A shared physical environment allows the use of specific spatial cues such as referencing an object or orienting the body or gaze towards a space of interest to help establish common ground and joint attention [7].

One of the prominent qualities that makes robots unique as a technology material is their embodiment in physical space. Robots can not only manipulate objects in space, but also express complex interactions through motion [28, 82, 1]. In Human-Robot Interaction (HRI), social robots have used motion to support people in a variety of related communication contexts through nonverbal behavior. For example, researchers have used social robot gaze to facilitate turn-taking [2, 106], and robot body gestures to mediate interpersonal conflict [68, 150]. Prior work has explored the use of *peripheral* interactive objects as communication supporters in a range of domains. One of the defining features of peripheral interfaces is that they exist primarily in the background and are only called to the foreground when necessary—they are not themselves interaction partners [72, 98]. In HRI, the notion of peripheral robots has also been explored. For example, a peripheral robotic device that draws attention when the conversation between a romantic couple becomes aggressive [68], one that shapes conversational dynamics in groups [150] or accompanies text messaging [111] have been introduced. Peripheral and embodied devices have also been suggested as *socially assistive* robots. Minimally expressive anthropomorphic robots have been used to support autistic children's communication [126] and to facilitate interaction between people with dementia and caregivers [105].

Social robots have the quality of being able to take on a specific social role in a conversation, such as an over-hearer [83] or a facilitator [101]. Prior work has suggested a metaphor of *sidekicks* as a way to think about social robots [97] that support a "protagonist" by existing in the periphery of attention and appearing only when needed.

In chapter 4, we leverage robots' advantages of embodiment and motion for nonverbal communication in the domain of AAC, to explore how social robots could be designed to support ACs' nonverbal communication. We use the metaphor of "assistive sidekick" to explore robot designs that support the AC, rather than function as independent agents. Using specific design parameters uncovered, in collaboration with an AAC user, Mark, we built a functioning sidekick and evaluate its use over two months. We describe this studies outcomes in Chapter 5.

2.3.2 Context Support through Language Technologies

Innovations related to language technologies in AAC systems include the use of word completion and word prediction, optimized dynamic keyboard designs, and the use of contextually aware sentence and word generation. More recently there has been growing interest to additionally lever-

age large language models (LLMs) for phrase generation and abbreviation completion. Natural Language Processing (NLP) strategies have been used in prior work to support ACs in accessing relevant utterances and increasing their speaking rate [64].

Word prediction

The most widespread application of NLP in AAC commercial systems is the use of word completion and word prediction [64]. Statistical language models are used to rank words to present them as possible candidates in current commercial AAC devices. Simple n-gram models are preferred as they require less computational effort. N-gram models use the current letter sequence to predict the most likely words that go next. Statistical models based on parts-of-speech are also used to determine the most likely continuing words in a phrase.

Optimized Keyboards

Keyboards with optimized layouts have been designed to decrease fatigue by minimizing user input effort, input errors and increasing text input rates. Keyboard designs are tightly connected to the access method utilized by AAC users. For users who use scanning with a binary input switch, researchers have proposed scanning techniques that blend statistical models with grid displays that dynamically keep track of the user’s selection to minimize input [124]. Researchers have also proposed dynamic keyboards with word and letter prediction that allow users to directly type on a keyboard that updates itself, select predicted words, and use letter completion [132]. For users who might not be able to select more than a few options due to screen real-state or motor constraints designs have placed multiple characters on each key (i.e., a keypad with multiple letter options per key [abc] [def]) and then used n-gram techniques to establish word probabilities used for disambiguation.

Encoding techniques

Encoding techniques are used in high-tech AAC systems to improve speaking rates and facilitate access to a large vocabulary. Encoding methods are language representations that use semantic compaction, that is using specific symbol combinations to generate multiple words. For example, a symbol for “juice” in combination with the symbol representing “verb” will indicate drink, while “juice” plus “adjective” will indicate thirsty [6]. Other approaches known as Utterance-based systems (UBS), link pragmatic features such as whether the user is talking about the past, present, or future with a user’s goal and have demonstrated to increase a user’s speaking rate. UBS has proven to increase users speaking rates by providing predictions according to the time perspective the user chooses to speak in [153, 152]. Some results of our workshop also indicate that attending to other pragmatic features such as identifying the communicative action of a user’s intended utterance may improve word prediction algorithms.

2.3.3 Emerging Language Technologies for AAC

Recent advances in large language models (LLMs) point to a great opportunity to leverage LLMs to support AAC communication. Nonetheless, we do not know how, in practice, AAC users could benefit from these enhanced use of language models and how they would like to interact with the models (give information, make selections in an interface) and how they would use them. In this study, we showed design concepts that showcased a system using partner speech to have conversations with participants about this idea, its usefulness and concerns around it.

Large Language Models (LLMs) are machine learning algorithms that can recognize, generate, and transform human languages by having learned patterns from large text-based data sets. Recent LLMs such as GPT-3 [25] have proven to learn from examples or text-based instructions known

as prompts [74] to generate language in context. By carefully crafting the inputs given to the model people can directly influence the output of a LLM by defining a desired task. LLMs have previously been used to support accessibility use cases including generating speech for AAC users [138, 27], and providing writing support for people with dyslexia [56].

This capability of understanding language to produce language makes LLMs a great resource that could enable AAC users to produce responses that are detailed and grammatically correct by only inputting a few words. Given a specific text-based instruction or example also known as prompt, LLMs can return plausible continuation or response to the given prompt. For example, *prompt: give me some fruits that start with "A", model: Apple, Apricots, Ananas*. There are many advantages to using prompts to retrieve customized output from a language model. Prompting does not require pre-training or fine-tuning of a model which can be expensive and require access to large amounts of data, which is a limitation in the AAC research field. Utilizing prompts as a prototyping tool can enable quick explorations on the types of input and outputs needed for the model to be most useful [74].

In chapter 7, we explore how prompting could support AAC users in their communication by presenting them with different pre-made prompts that can be configured at different levels: the type of context it uses, the type of input it requires from the user, and the task at hand (e.g., add details to a reply, share background information, turn words into requests).

2.3.4 Augmenting Conversational Context Awareness in AAC

In addition to improving performance using linguistic predictions, context-aware AAC solutions have been proposed to leverage context sources such as location [78], conversation partner word completions [46], and more recently the use of large-language models that utilize additional context such as a user's persona, keywords, and conversational context to suggest relevant content to an AAC user [138]. Using general dialogue data, some studies have shown substantial improvements in word prediction when language models consider the partner's speech in their prediction [160, 138].

Being able to reply to a topic while is still relevant requires good timing something that is challenging in AAC. Speaking rates for augmented communicators are slower than speakers who use their voice and for this reason often ACs' comments are often over a minute delayed making them out-of-context and hard to interpret [125]. In chapter 3, I found that fast changing topics produces missed opportunities for ACs to participate as they decide to erase comments after other conversation participants have changed the conversation topic [157]. Conversation partners can introduce new topics when talking to a third party while an AC is typing. Nonetheless, close conversation partners (CCPs) can also help in contextualizing a comment made by an AC after a topic has passed by asking follow-up questions to the AC that can help clarify their intended message. We identified these clarifying questions a CCP made to an AC as types of rituals that could support conversational agency by facilitating understanding at a lower cost to the AC [157]. CCPs also increased ACs participation through rituals involving asking ACs permission to answer a question, asking the AC to clarify information, and orienting their bodies towards the AC to ensure their participation.

Incorporating rituals into the design of AAC interfaces could be done by using NLP methods that can monitor the content of a conversation and enable different types of support as needs arise. For example having a list of interactive talking points to guide unfamiliar partners to learn how to better ask clarifying questions to an AC instead of guessing what they meant to say could better support AC agency in conversation. Additionally, the problem of talking "out-of-context" is tied to not being able to find the relevant reference the AC is addressing. The relevance constraint states that a new contribution to a conversation will be interpreted as a response to the most recent prior contribution. In AAC new contributions are not always connected to the most recent utterances. NLP techniques could augment AAC interactions by facilitating different ways in which

conversation participants could find the relevant context an AC is speaking to. This context support could take many forms such as audio excerpts, visuals showing key noun phrases, or tracking agency metrics automatically in a conversation in order to inform participants to be mindful of AC's speaking turns. NLP techniques could also help a system identify the most relevant context and suggest it to either the partner, to all speakers or to the AC to remind others about the intended context of an AC's utterance.

Using contextual information derived from a user's location, discourse genre, conversational topic, time of day and information from their conversational partner can be a powerful tool to improve expressive communication in AAC. In 2008, Wisenburn et al. designed and tested *Converser*, a program that used speaking partner's speech to predict contextually appropriate utterances for an AC [166]. *Converser* used speech recognition and a parser that posted noun phrases on an optimized AAC keyboard. The authors tested the system with dyads of non-augmented communicators [167] but did not find an improvement of speaking rate for augmented communicators. In 2018 Fried-Oken et al. continued exploring how word prediction could increase AC's speaking rate by leveraging a statistical language model, the communicator's recent vocabulary, and content suggestions from their conversational partner via a partner application named *SmartPredict* [49]. *SmartPredict* made it easier for ACs to access desired information but there is no knowledge as to how the application supports conversational agency. Research has explored using location relevant words to suggest vocabulary by leveraging location tracking systems [26], camera information [78], photographs [48], or geographically tagged language in a personal device [112]. NLP could further enhance these systems by searching, retrieving and displaying the most relevant words according to additional contextual information such as the conversational topic.

The type of discourse (e.g., description, narrative, expository, etc), the topic of a conversation, and even the dialogue acts (e.g., statement, opinion, agreement) can be identified by syntactic, grammatical, and semantic features in a text to support AAC-mediated communication. Topic-informed predictions have shown to save keystrokes in machine simulations of AAC systems [84]. Complementing topic-based predictions with web-based searchers using internet crawlers has also demonstrated keystroke savings for AAC systems [89]. Most recently, NLP researchers have been interested in exploring dialogue act classification for AAC [115]. Dialogue act classification involves identifying specific parts of speech such as "question", "opinion", or "open-ended statement" in a conversation. Dialogue act classification could enable smart phrase recommendations that could support ACs in speaking to a conversation topic in a timely manner, thus addressing the relevance constraint. Early explorations to develop dialogue act classification models for AAC have encountered challenges related to a lack of AAC conversation datasets and appropriately annotated data tags.

In Chapter 8 we introduce COMPA, a system that incorporates different types of contextual grounding strategies such as highlighting and alerting a communication partner about the current conversation portion an AAC user might be composing a message about. COMPA also integrates a large language models that provides starter phrases as suggestions that can help re-introduce a past topic into conversation.

2.4 Design Frameworks in Accessibility Research

Frameworks can provide a supporting structure for thinking about and doing research [96]. Frameworks in HCI and accessibility research often draw from theory to expand and propose new lenses through which we can approach research and design.

An influential frame, Value Sensitive Design, draws on moral epistemology to propose a principled approach to design that centers and maintains human values related to welfare, rights, and justice [51]. Value Sensitive Design argues that both social and the technical aspects of technology need to be considered. This dissertation builds on Value Sensitive Design's commitment to ground the design process of technology through conceptual, empirical, and technical investigations with

a commitment to human values, specifically agency.

In accessibility research, recent frameworks draw from critical disability studies to better inform how assistive technologies can center the lived experiences of people with disabilities [100]. Mankoff et al. make a case for how the field of assistive technology can be reframed in the light of disability studies by enabling developers to not focus only on the “technical merit of solving observable problems’ but to reflect and consider “the interplay of issues at work and acknowledge that there may not even be a “right” problem to tackle or a “right” approach to take.

The interdependence frame [9] draws from Disability Studies and popular media contributed by contemporary disability justice activists to expand influential approaches to assistive technology design such as Ability-Based Design [168]. While Ability-Based Design calls for adaptive user interfaces, taking an Interdependence lens highlighted the need to include considerations of the relationships people have with their disabilities, their devices, and other people. Similarly, by taking perspectives from disability studies into design practice, prior work has highlighted the importance of understanding disabled users’ stories and rich accounts to recognize the knowledge and labor disabled users already have expended [10].

The Agency in AAC design frame presented in this dissertation takes inspiration on the Interdependence frame and lays its foundation on literature from sociology which describes agency as being socially distributed [42]. In this way, this dissertation approaches the design process of accessible technologies with a socio-technical perspective. Additionally, the agency in AAC framework presented in this dissertation is based on Conversational Analysis (CA), the micro-analytic study of social interaction [58, 131]. CA describes and studies how conversation progresses, and this dissertation utilizes these methods to understand how AAC and non-AAC users carry out their conversation and how they structure it according to the systematic rules of turn-taking and other social and contextual factors.

Part I

Agency as a Design Framework

Chapter 3

Conversational Agency in AAC

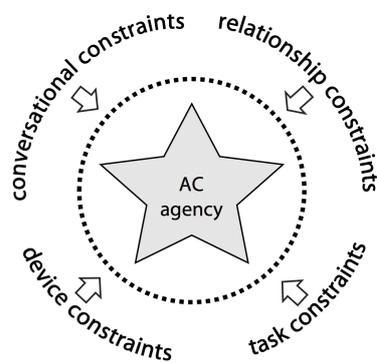


Figure 3.1: Conversational constraints (*e.g.*, who can speak, who can be addressed), the task at hand, the relationship with other speakers, and the device being used can all impact the expression of an augmented communicator’s (AC) agency in conversation.

3.1 Overview

As a first step towards making communication possible and efficient, designers and researchers of AAC have focused on improving device throughput, for example via text prediction [167], and vocabulary sorting [5, 78], and improving expressivity, with awareness displays [146] and customizable voices [45, 104, 121]. Research from communication science has focused on increasing communicative competence of ACs [90, 92]. Other researchers have recognized social factors including the role of communication partners in how an AC communicates [46]. Recent AAC research has started to take a critical view of how AAC technology mediates communication [70], but there is still no model of how constraints specific to AAC impact the expression of agency in AAC in-person interactions. Prior work has not yet determined how different types of conversational partners impact an AC’s agency.

In this work, we introduce using conversational agency as a frame to study AAC user’s experiences in conversation. We specifically consider Gibson’s definition from sociology [54], which explains that individuals can exercise agency when conversational constraints loosen enough for them to contribute to the conversation. These conversational constraints include when a person can speak, whom they can address, what can be said, and what can be expected from others by way of cooperation. The four conversational constraints are summarized in Table 3.1). Gibson also distinguishes two types of agency. *Technical agency*, is exercised when individuals contribute to the conversation (Figure 3.2). As individuals express technical agency by participating they can then advance their objectives in conversation, which he identifies as *colloquial agency*.

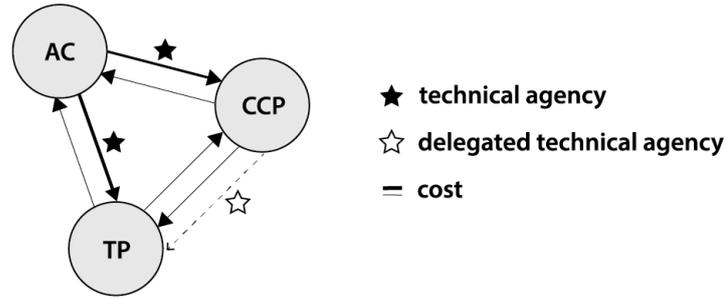


Figure 3.2: Conversational, task, relationship and device constraints can change the pathways for communication (arrows) between AC, CCPs and a third party (TP) and the effort, or cost, required to communicate along the pathways (arrow thickness). For example, time-sensitive conversational constraints (*e.g.*, replying fast while it is still relevant) can encourage communication along the lowest cost paths (lower width arrows) encouraging communication via their CCP rather than direct AC to TP communication. When this delegated technical agency enables the realization of goals for ACs we call it Colloquial Agency.

Constraints	Governs	Contributing factors
One speaker	Current speaker	Time to start speaking
Participation shift	Next speaker	Addressing next speaker (<i>e.g.</i> by gaze, name, topic)
Relevance	Topic of contribution	Time spent speaking without pause
Ritual	Type of contribution	Previous speech act (<i>e.g.</i> question)

Table 3.1: We summarize the four conversational constraints that impact agency according to Gibson, what governs the constraint, and factors that contribute to the constraint application.

In addition to conversational constraints, we also consider constraints that are particularly relevant in the case of AAC communication and task-oriented dialogue (Figure 3.1): relationship with a communication partner (*e.g.*, years knowing each other, comfort, previous communication experience), device properties (*e.g.*, throughput, input type, vocabulary, feedback type), and task constraints (*e.g.*, communication channel available, output type, information owner). We are interested in understanding how ACs exercise conversational agency in the face of unique constraints such as the device they use, their relationship with the people they are talking with, the task at hand and the general conversational constraints explained by Gibson (Figure 3.1).

To address this, we carried out an observational study with ACs and their close conversational partners (CCPs). CCPs such as parents and paid aides, who have caregiving roles, often support ACs by facilitating conversations with others [15]. We consider scenarios in which there is an unfamiliar third party (TP) to study how CCPs facilitate the conversation between the AC and the third party, and how this impacts the ACs’ agency. Our findings show that CCPs support ACs’ agency by supporting communication with third parties. CCPs also increase the amount of information exchanged on behalf of the AC at a lower cost to the AC through explanations and by breaking down complex questions into yes/no questions. But, such explanations and narrow questions can lead to a reduced set of AC responses, and missed opportunities for AC to participate. In this paper, we contribute an empirical understanding of agency in AAC and conclude with ideas on how future designs can consider CCP participation while facilitating ACs’ to exercise their agency in conversation.

3.2 Method

To study the participation of close conversational partners (CCPs) to understand how to facilitate augmented communicators (ACs) to exercise their agency we address the following research questions: **RQ1**: How do conversational, relationship, device and task constraints impact CCPs' participation? **RQ2**: How does CCPs' participation impact ACs' agency? and **RQ3**: When does technical agency impact colloquial agency?



Figure 3.3: Study setup. The CCP (left), the AC (middle) and the experimenter(right) sit at a round table. A scene camera captures this image, a camera behind the AC captures activity on the AAC device, and a camera mounted on the AC's wheelchair captures AC's facial expressions.

3.2.1 Tasks

We use three tasks to investigate how factors such as potential to speak (*e.g.*, who holds the knowledge in the conversation, and who are the active participants) and goals (*e.g.*, an answer to a question, versus a creative decision) impact ACs' technical and colloquial agency when working with a CCP. We include the participation of a third-party (TP), who is an unfamiliar conversation partner to the AC (experimenter) to analyze differences due to the relationship constraint (Figure 3.3). The TP has experience communicating with ACs but not the particular participants. The TP holds a relationship similar to the description of partners in the fourth circle (*e.g.*, workers who communicate with ACs as part of their profession) and the fifth circle (*e.g.*, people ACs have not met before) in Blackstone's Circles of Communication Partners [15].

Interview task

We conducted experimenter-led semi-structured interviews with each AC and CCP pair in order to 1) investigate AC technical agency in a structured conversation assisted by a CCP, and 2) gather background information about the AAC device and the AC-CCP pair. The experimenter asked all questions directly to the AC. Questions were about the device and the ACs' experience with it. These are topics in which the AC and CCP could potentially both know the answers, so they shared the same potential to respond (Figure 3.4, far left). The experimenter addresses all questions directly to the AC, to control the participation shift constraint and evaluate if this helped the AC secure turns and participate more. We expected the interview to encourage open conversation to answer RQ1.

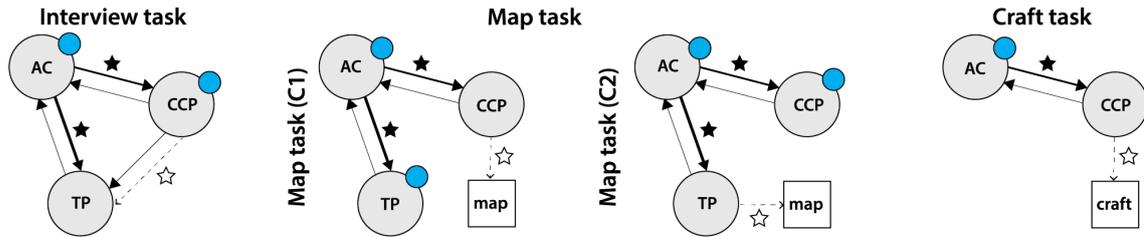


Figure 3.4: The interview task mirrors open conversation except that the third party only directly addresses the AC (missing arrow between the third party and CCP) and only the AC and CCP share knowledge to answer the interview questions (small blue dots). In the first map task condition (C1), the AC and the third party can see the correct route (small blue dots) and the AC must communicate to the CCP what to draw on the map. In the second map condition, the AC and CCP share the correct route and must tell the third party what to draw on the map. In the craft task, the AC generates ideas for their craft which is largely constructed by the CCP (the third party does not participate).

Map task

Inspired by prior work [66, 63] that used maps to examine how ACs and their partners achieve common ground, we investigated how CCP-AC partnerships differ from TP-AC partnerships when working together to complete a map. For this task the AC has access to a map displaying a route between landmarks and must direct another person (either CCP or TP) to draw the given route on their blank map.

In condition one (map C1), the TP acts as an assistive partner who can also see the map and the CCP is the follower who has the blank map (Figure 3.4, second from left). In map condition two (map C2), the CCP is the assistive partner and the TP is the follower (Figure 3.4, second from right). We use the map task to answer RQ1 and RQ2 and also explore RQ3 to understand what type of interactions contributed to the AC being able to successfully communicate the map's route (technical agency) and their partner drawing it correctly on the map (colloquial agency). Higginbotham *et al.* [66, 63] found that CCP often set rules or strategies to solve the map task in an efficient way, so we hypothesize the setting of rules or strategies as a potential action from CCPs that can impact ACs' expression of agency. We also expect to see different interaction patterns when there is mainly AC-to-CCP communication (condition 1), as opposed to AC-to-TP communication (condition 2).

Craft task

To further explore RQ3, we designed a craft task in which the AC and CCP collaborated to design an image collage (Figure 3.4, far right). The TP did not participate in this task. Similar to the map task, we designed a craft task to explore how AC and CCP collaboration supports the translation of ideas expressed by the AC (technical agency) to design decisions on the final collage artifact (colloquial agency). The craft task involves choices based on preference rather than a single correct route as in the map task. In this activity, the AC holds the knowledge of their creative choices, while the CCP can physically explore the materials available and physically creates the final artifact through delegated agency.

3.2.2 Operationalizing Agency

We used quantitative video coding analysis to develop both overall and task-specific metrics for technical agency, colloquial agency and CCP participation to address our research questions. We considered the amount of conversational contributions made by CCPs and ACs and classified them

with specific communicative function types as in [31, 91]. To address RQ3, we also examined the resulting artifacts of the map and craft task to evaluate how participation (technical agency) impacted the map or craft (colloquial agency).

To measure conversational contributions we first annotated the start and end of each participation event in the videos (e.g., the CCP speaks to give an explanation, the AC gestures to confirm). Then, each participation event was labelled with an ID (e.g., CCP, AC, third party), the form of the interaction (e.g., speaking, gesturing, vocalizing), and the type of communicative function (e.g., giving an explanation, confirming information). We selected communicative functions labels specific to our research questions (see supplemental material for full guide) and only labelled actions, gestures and vocalization that served a communicative function as in [31]. Two coders watched all of the videos. We determined inter-rater reliability from a sample of 25% of video data coded independently by two coders. We sampled each 0.5 second, and achieved a Cohen's Kappa of 0.87 overall communicative functions labels for all participation events. Videos were transcribed and time-marked to capture sequences of utterances and turns.

AC technical agency

We measured occurrences of AC's technical agency as individual contributions to the conversation through speech, gestures, and vocalizations. Through all tasks, we consider the length and number of conversational turns taken in each modality. In the interview, we separately considered technical agency (whether an AC gives an explanation or a yes/no answer to a question), and delegated technical agency (CCP answers the question with AC's permission). We do this by identifying specific AC communicative functions such as: requesting help (RH) or time (RT), giving an explanation (E), granting/denying permission (GP/DP), and confirming/denying information the CCP has said, other than a permission request (COI/NOI). In the map tasks, we considered the number of strategies explicitly stated by the CCP and the AC to solve the task. In the craft task, we considered the total number of ideas proposed by the CCP and the AC.

AC colloquial agency

We examined the resulting maps routes and collage elements as a proxy for colloquial agency (achieving individual objectives). For the map task, we considered the number of accurate and inaccurate landmarks on the final map. For the craft task, we considered each element in the collage, examining who proposed the element (e.g., CCP or AC) and who agreed to the idea of including the element (e.g., none, CCP, or AC).

CCP participation

For all tasks, we related CCP participation to the AC's participation (technical agency) by calculating similar metrics to those calculated for the AC: number of conversational turns, amount of speaking time per task, number of strategies stated in the map task and number of ideas contributed during craft task. In addition to this, we annotated specific types of CCP participation by communicative function: asking permission to explain (P), giving an explanation (E), lending assistance (A) such as conversational repair and physical assistance through actions, and asking clarifying questions (QC) to the AC.

3.2.3 Participants and data collection

We recruited 8 participants: 4 expert augmented communicators (ACs) with cerebral palsy and their corresponding close conversational partners (CCPs) (Table 3.2). We call ACs expert AAC users as they have been using AAC devices for more than 10 years, on a daily basis. In addition to their device, AC participants also used nonspeech vocalizations (e.g., grunting) and facial gestures

Table 3.2: Our eight study participants, four augmented communicators (ACs) and their corresponding close conversation partners (CCPs). Using our transcripts, we calculated the throughput or total words spoken over minutes typing (WPM) for each AC.

ID	Age	Gender	Years using AAC; current device	WPM	AAC Device Setup	Partner
AC1	23	M	19; 10	2.65	Tobii Dynavox Maestro; head-controlled switch, scanning access; phrase-level voice output	Mother (CCP1)
AC2	38	F	32; 1.5	8.94	Tobii Dynavox i-12; wheelchair joystick, direct access; phrase-level voice output	Father (CCP2)
AC3	50	M	35; 1.5	7.03	Accent 1000; head tracker, direct access; utterance-level voice output	Paid aide for 0.7 yrs (CCP3)
AC4	50	F	33; 23	22.24	DeltaTalker; head-mounted pointer, direct access; utterance-level voice output	Paid aide for 4 yrs (CCP4)

(*e.g.*, blinks, smiles and eye-contact) to communicate. All ACs had closed fists and high spasticity. AC1 and AC4 had hand restrains while AC2 and AC3 used arm gestures. We collected audio and video recordings from three video cameras to capture all tasks: 1) a camera placed at a distance to capture the entire scene including all participants, 2) a camera placed behind the AC to capture the AAC device’s screen, and 3) a small GoPro camera in front of the AC (either on the AC’s wheelchair mount or on the table in front of the AC) to capture facial expressions.

3.2.4 Procedure

An experimenter (the first author) first conducted semi-structured interviews with participant pairs, directing all questions to the AC with the CCP present. Interviews took between 35 and 55 minutes to complete including rest breaks. A complete list of interview questions and study materials is in the supplementary material. We then gave participants a limited amount of time ($\mu=12$ minutes) to complete each of the two map task conditions, and we counterbalanced the order of the conditions between participants. To select maps for each pair, We used two randomly selected maps from the HRCR map corpus¹. We edited these maps to include large font and minimal graphics, and provided them to the participants on paper. After the map task, participants spent 12-18 minutes completing the craft task. We provided participant pairs with materials to create a collage craft that

¹<http://groups.inf.ed.ac.uk/maptask/maptasknxt.html>

included construction paper, magazines, writing utensils, stickers, scissors, tape and glue. Pairs could make a collage either: 1) for a loved one to give as a gift, or 2) about a place they wanted to visit on vacation. We invited each pair to take home their collage. At the end of the study we asked each pair a series of debrief questions together, and then a different set of questions to each AC and CCP individually (without the other present). With both participants present, we asked about likes, dislikes and challenges encountered during study tasks. To the AC and the CCP individually, we asked about strategies and challenges experienced while communicating with their partner and with others outside of the study.

3.2.5 Study Limitations

Our study required in-depth qualitative work to observe behaviors and social conversation in a constrained environment. Despite potential observation effects, we observed meaningful differences across tasks on the way participants behaved. Our study investigates specific partner configurations as described in Figure 3.4. These are examples of possible partner configurations but do not encompass all possible conversation scenarios. There are still many other types of AC to CCP relationships that may impact agency and need to be investigated (*e.g.*, “friend-tendants” or colleagues). The inclusion of the TP who is familiar with AAC but did not know the participants previously resembles the scenario of an AC who meets a professional for the first time and is accompanied by their CCP.

3.3 Results: Constraints and Participation

Our first research question asked: *How do conversational, relationship, device, and task constraints impact CCP participation?* Conversational constraints regulated CCP participation by balancing turns and enabling topic dominance through long contributions. CCPs who were parents to the AC participated much more than paid aides across all tasks. ACs using devices with lower WPM experienced more CCP participation. The task goals and the access to the information needed to complete the task also impacted CCP participation.

3.3.1 Effect of Conversational Constraints on Participation

CCPs participated less when the third party directly addressed the AC in the interview task and participated more in subsequent turns (participation shift constraint). CCPs also balanced their own explanations by asking clarifying questions or asking for permission before answering or expanding on a question on behalf of the AC (ritual constraint) engaging in a collaborative sharing of information. Long turns by the CCPs, often during AC typing, impacted ACs’ opportunity to make relevant contributions (relevance constraint). CCP participation lowered when the AC took more turns and longer turns and used utterance-level voice output that allowed the AC to hold the floor while typing (one-speaker constraint).

Participation shift constraint

The participation shift constraint states that an individual has more opportunities to contribute when addressed directly by the current speaker. Conforming to this constraint, the AC spoke more often relative to the CCP when directly addressed by the third party (Figure 3.5). When CCPs participated in subsequent turns after the AC took the first turn, CCPs expanded on ACs’ responses by (1) giving an explanation of the ACs’ response to provide context (with or without permission), and (2) asking the AC a follow up question to achieve clarity. During such explanations, CCPs often expanded on ACs’ yes/no responses to the third party:

TP: Do you use your device for social media?
 AC2: No
 CCP2: Yes and no . . . We just don't do it at the moment . . .

The CCP most often responded first to the third party when the AC struggled to answer the question. For example, after one of the interview questions CCP2 asked for permission to respond on behalf of AC2 after noticing she was stuck on a page on her device and was taking a long time to answer the question. As the duration of the interview increased, the participation shift constraint loosened, participation shifted to and from all directions (CCP to TP, CCP to AC, AC to CCP), and the CCP spoke more often. The only exception was CCP4 who only contributed to clarify AC4's device pronunciation.

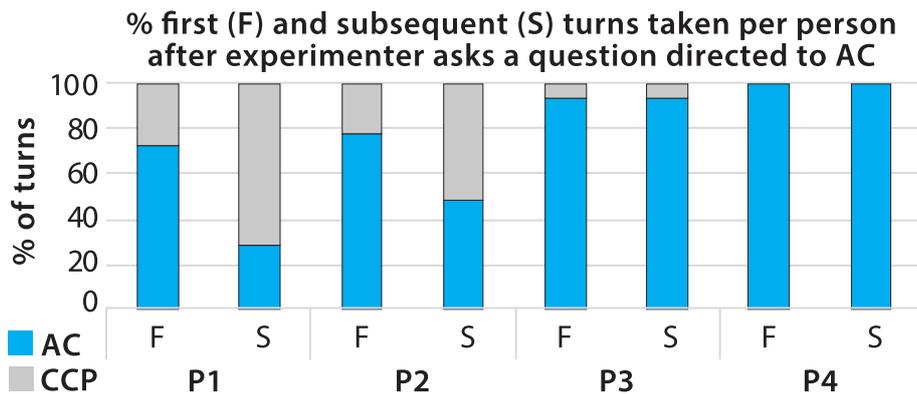


Figure 3.5: Overall, ACs contributed a higher percentage of turns immediately after they were addressed by the third party (F). CCPs in pairs 1 and 2, contributed a higher percentage of the time in subsequent turns (S).

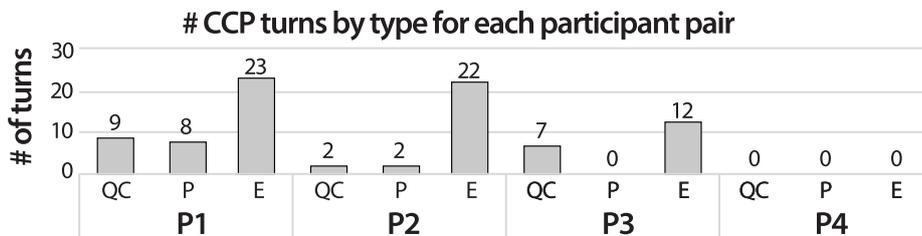


Figure 3.6: When CCPs took a turn during the interview, they most often gave explanations (E). CCPs also asked questions for confirmation by the AC (QC) and asked permission before providing explanations (P).

Ritual constraint

The ritual constraint suggests that participants regulate the conversation to maintain a positive social standing (*e.g.*, by involving a new speaker, answering a speaker's question, or filling silence). CCPs upheld the ritual constraint by providing direct explanations (E) which served to answer questions and fill silences. They also upheld the ritual constraint by asking permission before giving an explanation (P) and by asking clarifying questions (QC), which served to involve the AC (Figure 3.6). ACs reciprocated clarifying questions with a simple confirmation or denial of the information if they were the next to speak:

TP: Is there any other device that you use for communication in general or for computer access?
AC1: Yes.
CCP1: To use the device as a computer, is that what you are thinking?
AC1: [*Blinks once to say "no."*]
CCP1: ... Or are you thinking about the iPad? The one I am thinking of ...
AC1: Low-tech.
CCP1: Oh there is low-tech. Oh my gosh, thank you!

When CCPs asked for permission, they did not necessarily wait for permission to be granted by the AC. For instance, AC1 only granted permission on 4 of the 8 occasions when CCP1 asked (Figures 3.6 and 3.7). CCPs also provided a portion of their explanations while the AC was typing (CCP1, CCP2, and CCP3 gave 50%, 47% and 12% of their explanations during typing respectively; CCP4 did not give any explanation).

% AC interview turns for each interaction type by participant pair

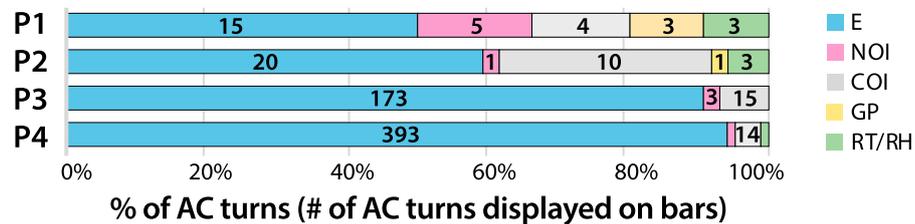


Figure 3.7: When ACs took a turn during the interview, they most often gave explanations (E). Other turns were spent confirming (COI) or denying (NOI) information given by the CCP, granting permission (GP) to CCPs to provide explanation, requesting help (RH) and requesting time (RT). ACs in P1 and P2 explained in less than 60% of turns, while ACs in P3 and P4 explained in more than 90% of turns.

Relevance constraint

As CCP participation increased in subsequent turns, conversation topics moved quickly and the time window ACs had to respond shrank, tightening the relevance constraint. In consequence, when ACs finished typing something they wanted to say, they had to choose to either erase it, or to say it even if this required making an off-topic comment. For example, on two occasions, AC1 typed a relevant answer but then erased it as the conversation continued and the comment lost relevance, all while CCP1 was speaking. The first time AC1 erased his contribution, the CCP was giving a long explanation (1.8 minutes). The second time, AC1 erased a contribution after the third party asked a second question and CCP1 replied to it with an explanation that took about a minute:

TP: For how long have you been using the Series 5?
CCP1: Yeah for how long... Do you know?
AC1: [*Gestures no by blinking once, begins typing.*]
CCP1: I am going to say like ten years ... he is actually been using a device for 19 years ... [*continues explaining for a long time*]
TP: Do you use your device every day?
AC1: [*Erases message typed, unintelligible.*] Can you hold on a minute, please?
TP: Yes.
AC1: Every day.

On two other occasions, AC1 and AC2 contributed off-topic comments to the conversation. In one instance, CCP1 was trying to remember the name of old hardware that AC1 had used. CCP1 eventually changed the topic. After nearly two minutes, AC1 spoke the hardware's name. CCP1 was taken by surprise. The second off-topic comment also occurred after a long explanation by CCP2, that had caused the topic to move on. In contrast, CCP3 and CCP4 spent less time explaining and offered explanations less often, and neither AC3 or AC4 experienced off-topic or erased comments.

One speaker constraint

Talking over ACs' speech utterances as they are being formed is a violation of the one speaker constraint. Participants with utterance-level speech (AC3, AC4) engaged in more open-ended participation, providing more explanations during the interview (Figure 3.7) than participants with phrase-level speech (AC1, AC2). AC3 and AC4 were able to participate more (Figure 3.8) and were less likely to be interrupted.

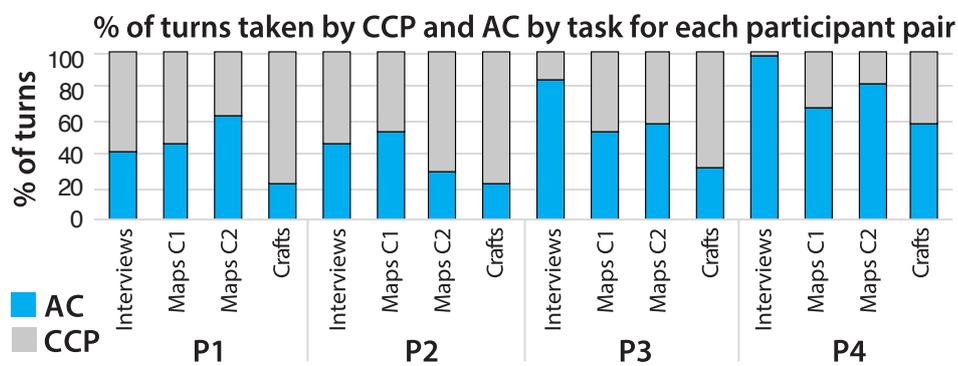


Figure 3.8: Percentage of turns taken by the AC and CCP per task for each participant pair.

3.3.2 Effect of Relationship Constraints on Participation

CCPs who are parents to the AC and therefore lifelong CCPs (CCP1, CCP2) participated more often across all tasks (Figure 3.8). In the interview, parents took speaking turns to provide explanations and ask for permission before explaining (Figure 3.6), while paid aides (CCP3, CCP4) participated primarily to clarify ACs' contributions and speech. In the map task, lifelong partners and aides participated relatively equally. In the crafts task, lifelong partners participated more and contributed more craft ideas (Table 3.3), while paid aides took comparatively fewer turns relative to ACs.

Across all tasks, ACs interacting with lifelong CCPs also used gestures and vocalizations to communicate more often than ACs interacting with paid aides (Figure 3.9). In one instance, during an explanation about word pronunciation by CCP1, AC1 communicated with a glance to the CCP. In response, CCP1 immediately chuckled, saying "*alright, I will stop talking now*". In another instance, when the third party asked AC2 an interview question, AC2 extended her hands to her father, CCP2, who immediately responded by holding her hands and answering the question. Unlike paid aides, lifelong CCPs occasionally interacted directly with the AAC device by directly operating the device to complete a word (CCP1) or moving the device to one side to allow the AC more room to color and choose materials in the craft task (CCP2).

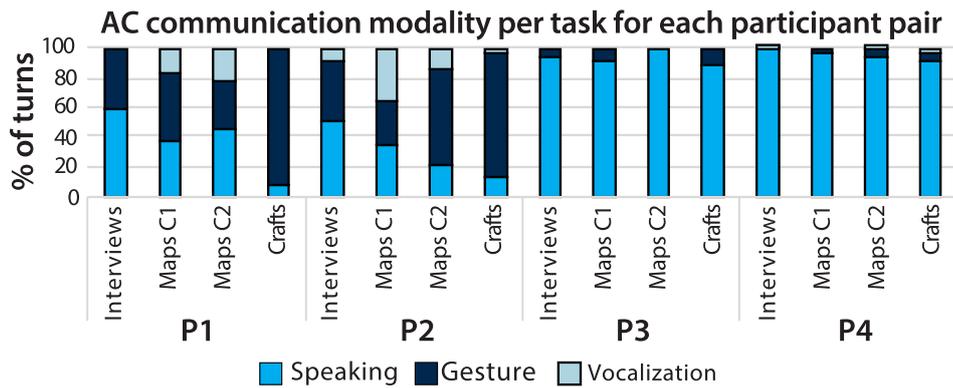


Figure 3.9: Modes of communication used by AC per task

3.3.3 Effect of Task Constraints on Participation

Access to information needed to complete each task, the task’s goals, and the communication partners’ roles all impacted CCP participation. All participants completed all tasks except for pair 2, who did not complete either map condition by decision of AC2, who found the task very challenging.

Access to task-relevant information

CCPs contributed substantive explanations and ideas when they had access to information relevant for the task. In the interview task, CCP1 and CCP2 (lifelong CCPs) contributed more explanations based on their knowledge of the AC’s experiences (Figure 3.6). In the craft task, CCP1 and CCP2 knew the recipients of their collages, and as a result suggested more ideas than CCP3 and CCP4 (Table 3.3).

In contrast, when ACs uniquely held the information required to complete the task, like in the map C1 and craft tasks (the route to the map and their creative ideas, respectively), all CCPs, except CCP2, participated more often by actively retrieving information from ACs (Figure 3.8). In particular, CCPs asked yes/no, short answer, and follow-up questions to try and retrieve information quickly (e.g., “what is the first letter of the next location?”, “what should we do?” and “how about this picture of the beach?”).

Task goals

Under the pressure of the time-constrained, goal-oriented map tasks, CCPs participated less often than in the more time-flexible, open-ended craft task (Figure 3.8). In the map task, CCPs proposed many strategies and quickly settled on a strategy once they found a successful approach (CCP1, CCP2, CCP3, and CCP4 proposed 7, 4, 3, and 9 strategies respectively across both conditions). ACs participated by stating directions in the map task, but ACs infrequently suggested strategies (only AC2 and AC4 proposed one strategy each over both conditions). Although CCP suggested strategies intended to make the task easier (“*you don’t have to give me the whole word ... just give me a couple letters*”), the strategies limited the ways that ACs could carry out the task. For instance, CCP1 suggested AC1 type the first letter of each landmark and CCP1 would type the rest. But when AC1 typed the letter “f,” CCP1 mistakenly completed the word “finish” twice, even though AC1 tried erasing and spelling another word, “Flag ship,” which was the last landmark on the route. The strategy set by CCP1 prevailed and it was challenging for AC1 to correct it. In the collaborative craft task, CCPs took more turns relative to ACs to suggest multiple ideas for the crafts, allowing ACs more options before settling on a solution.

Communication Partners' roles

CCPs provided assistance more frequently when performing tasks that required primarily AC-to-TP communication compared to when performing tasks that required primarily AC-to-CCP communication (Figure 3.4). For the interview and map C2 task, which mainly required AC-to-TP communication, the CCP frequently assisted and repaired communication between the AC and the third party by clarifying device errors and pronunciations (in the interview) and groundings (in the map). During the interview, CCPs provided 18, 17, 11, and 3 clarifications respectively. During map C2, CCPs grounded the conversation by questioning the TP to check if they understood the AC's communication (e.g., "do you have that landmark?", "yeah, he blinked twice, that is a no," "she is typing"). In the craft and map C1 tasks, which required AC-to-CCP communication, ACs used gestures and vocalizations more often than in tasks that required AC-to-TP communication (Figure 3.9).

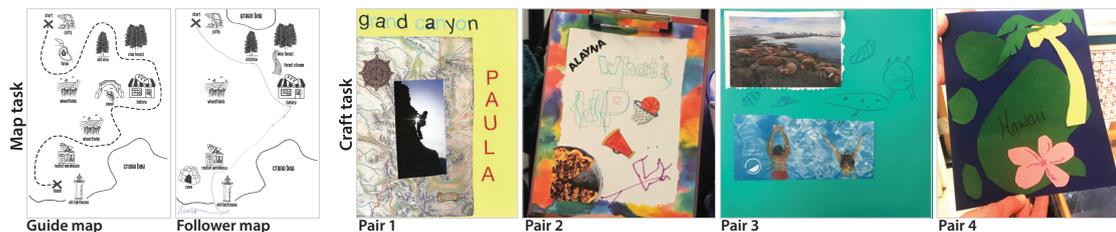


Figure 3.10: Example materials for the map and craft task. For the map task, we show an example guide map with route information (visible by AC) and follower map (visible by CCP or third party) that shows only the starting point, requiring the follower to draw the rest of the information (displayed in pen). For the craft task, we show the collages produced by each participant pair. Names that appear on collages are the intended receivers.

3.3.4 Effect of Device Constraints on Participation

Properties of the AAC device (e.g., throughput and input type) impacted AC and CCP participation. AC1, AC2, and AC3, who experienced lower device throughput (Table 3.2), displayed high ratios of CCP-to-AC participation (Figure 3.8) compared to AC4 who experienced the highest device throughput. While CCP1, CCP2, and CCP3 gave explanations in the interview (Figure 3.6), CCP4 participated only 3 times to clarify device pronunciation. AC4 also managed the conversation (e.g., "oh, let me think about that," "did I answer your question?") and elaborated more on her answers compared to other ACs. On the other hand, CCP1 facilitated abbreviated communication by AC1, who had the lowest WPM, by requesting that the AC type individual letters instead of full words in the map task.

3.4 Results: Participation and Technical Agency

Our second research question asked: *How does CCPs' participation impact ACs' technical agency?* We identified that CCP participation: 1) supported AC contributions by creating multi-modal communication channels, 2) increased information exchanged at a lower cost via delegated agency, but also 3) reduced AC contributions through missed opportunities.

3.4.1 Supporting Independent AC's Communication

CCPs facilitated communication with third parties by knowing how to interpret their AC partner's unique communication characteristics. For example, CCPs assisted in repairing misunderstand-

ings by clarifying AAC device pronunciation, informed the third party about the AC's status when there was a device error, and let the third party know what AC gestures or nonverbal behaviors meant. CCP attributes (*e.g.*, previous experience, relationship) enabled this process. Thus, the presence of a close conversation partner opens new communication paths for the AC and extends their communication possibilities beyond the AAC device. Physically, CCPs also enable otherwise challenging tasks (*e.g.*, bringing maps closer for ACs to see, or moving crafting materials closer for ACs to draw with) to support independent AC contributions. Outside of the study, CCPs can encourage third parties to elicit AC contributions, as mentioned during the debrief interviews:

“One of the things I do to help with that is by turning my body towards him, so that they are not relating to me but they are relating to him. I am helping orient the conversation towards him.” – CCP1

3.4.2 Increasing Information Exchanged

CCPs can increase information being shared in a conversation with a third party by expanding on information previously provided by the AC with permission from the AC (delegated technical agency) or by directly answering questions and confirming the information with the AC afterwards:

“I would ask him, would you like me to explain that more? . . . but if he would like to say it himself he could continue but I will usually will help him when I know there is more to something.” – CCP1

The latter strategy facilitates confirmation or correction of the information by the AC, but it may leave less room for independent AC explanations as there is no time allowed for the AC to directly answer or frame their response as they desire. ACs confirm that outside of the study, CCPs often anticipate what the AC will say by guessing (AC1, AC3, AC4), or by looking directly at the device (AC4). But at the same time, one AC reported that such guessing is “*useful 50% of the time*” and AC4 reported that while she does not prefer over-the-shoulder reading it can be occasionally useful:

“[At] church, it's often times noisy so in that case I don't mind if a person looks on my display. . . . I would rather be understood than them hearing me.” – AC4

AC4 mentioned that whether she liked other people guessing varied depending on how she was feeling:

“It varies. If I am on edge I do not appreciate people anticipating my words. I can and I want to articulate myself and my way. But that isn't me very much. I am pretty easy going.” - AC4

When ACs were asked for ideas of what CCPs could do instead of guessing, AC1 explained: “*I am not sure - because they can't read my mind. But if it is something I usually need, it is useful.*” AC3 said he preferred guessing when using headspelling, in which his CCP would guess as he spelled letters using his head.

3.4.3 Creating Missed Opportunities

Though CCPs aim to increase information exchange and lower the cost of expression for ACs, their actions can actually reduce ACs' technical agency by creating “missed opportunities” to contribute. For example, ACs occasionally erased what they started to type or were required to provide short answers. CCPs' lengthy explanations to interview questions introduce new topics and information that tighten the relevance constraint and reduced the time ACs have to contribute.

In addition, “guesses” that are phrased as questions meant to anticipate what the AC is thinking can facilitate information exchange at a lower cost (through a yes/no confirmation) if the CCP’s guess is right, but can also reduce ACs’ technical agency by limiting their contribution to short answers. These guesses can introduce new ideas that can be hard to repair if they differ from what the AC really wanted to say.

3.5 Results: From Technical to Colloquial Agency

Our third research question asked: *When does technical agency impact colloquial agency?* In other words, how does participating in conversation translate to successfully advancing goals? We found that collaboration and agreement between partners is necessary for technical agency to translate into colloquial agency.

In the map task, ACs could express technical agency by selecting a strategy to describe the map route and by contributing detailed directions. Among ACs, only AC4 expressed technical agency by deciding on the strategy:

- CCP4:** I say divide the paper left and right . . .
TP: I see the start point on my map, where should I go from there?
AC4: [CCP4], I think you are confusing [experimenter].
AC4: I would go to the g-r-a-v-e-y-a-r-d.

Additionally, AC4 would not only say the landmark’s name (as other ACs did) but also followed up to ask where the landmark was located with respect to others. As a result, pair 4 had no inaccurate landmarks, a successful translation of technical agency to colloquial agency.

In pairs other than P4, AC technical and colloquial agency were limited. In these cases, CCPs proposed the strategy used to complete the final map, reducing ACs’ technical agency. Their final maps contained inaccurate landmarks, mostly due to misunderstandings of what landmark the AC was referring to, which was the result of errors in grounding or mutual understanding. These inaccurate landmarks represent a failure of AC colloquial agency, because they were unable to advance their objectives (identifying landmarks) in the task.

Table 3.3: Collage elements in the final craft that were proposed by the AC, proposed by the CCP and agreed to by the AC, or proposed by the CCP and included in the artifact without explicit agreement from the AC.

	P1	P2	P3	P4
Total # of ideas proposed by AC	1	6	4	4
Total # of ideas proposed by CCP	16	16	4	10
# collage elements proposed by AC	1	6	4	4
# collage elements proposed by CCP with AC agreement	4	4	3	3
# collage elements proposed by CCP without AC agreement	2	0	1	1

In contrast, in the craft task, ACs achieved colloquial agency by providing ideas that made it into the craft. All of the ideas proposed by ACs appeared on the final collages, while only 37.5%, 25%, 80% and 40% of ideas proposed by CCP1, CCP2, CCP3 and CCP4 (respectively) appeared on the final collage. More CCP suggestions did not lead to more CCP ideas on the final craft. In fact, an idea rarely appeared on the final craft without some form of AC input—93% of CCP ideas on final crafts were first confirmed with AC partners (Table 3.3). CCP1 and CCP2

asked yes/no questions to confirm all decisions, while CCP3 asked open-ended questions to get ideas from AC3 (e.g., “what is on the beach?”). Pair 4 discussed ideas instead of only confirming CCP4’s suggestions:

ccp4: Would you like to do a pineapple? . . .

ac4: Thinking-of-a-comma-l-e-i. I was thinking of a lei.

ccp4: A lei, okay. There is pink around somewhere. [*browses crafting materials*].

In summary, AC participation impacted the resulting artifacts. The translation from participating (“technical agency”) to goal realization (“colloquial agency”) was contingent upon cooperation with CCPs, and made possible through interdependent effort, such as having CCPs carry out ideas suggested by ACs.

3.6 Discussion and Reflection

Our research confirms agency as a useful frame to consider social interactions of augmented communicators (ACs) in conversation. Constraints of conversation, relationship, task and device create opportunities and challenges for ACs to express agency. Building on our increased understanding of AAC interactions beyond the device, we suggest opportunities for future research to further understand and promote AC agency under existing constraints.

3.6.1 Designing for Conversational Constraints

Conversational constraints are influenced by other speakers, including close conversational partners (CCPs) or unfamiliar third parties (TPs). Our findings suggest that enabling AC’s to have more control over conversational constraints may lead to more equal participation in conversation.

Participation shift constraint: In the interview task of our study, ACs participated more frequently when directly addressed with an interview question than in subsequent turns. Addressing ACs more frequently in conversation can increase AC participation, and future work could consider how to ensure speakers are addressed equally through technology or conversational methods. For example, prior work aimed at balancing group speaking participation used a rotating robotic microphone that oriented to each speaker in turn [150]. Such work could be extended to balance group communication with ACs by developing social robots that encourage turn taking strategies for speakers with varying speeds (e.g., moving to other speakers while the AC types, and returning to the AC when they are ready to speak).

Ritual constraint: CCPs increased ACs participation through rituals such as asking ACs permission to answer a question, asking the AC to clarify information, and orienting their bodies towards the AC to ensure their participation. Future work may consider methods and technology that monitor conversation dynamics and leverage rituals to support AC participation when needed. Prior work that monitors dominant speakers and uses gaze of a social robot to guide attention to less dominant speakers [107] could be extended to increase AC agency when necessary. We could also consider methods to teach aspects of CCP rituals to unfamiliar TPs, such as a list of interactive talking points to guide TPs to better support AC agency in conversation.

Relevance Constraint: An AC is disproportionately impacted by the relevance constraint as an AC has to first type a message before sharing it. Once the AC finishes typing, it might be too late as the topic of the conversation has shifted. Prior work in AAC uses contextual information from the outside world to suggest words and phrases related to the current visual scene [78]. Using additional contextual information from the conversation such as audio, visuals and information on who is speaking could help the AC quickly reply to the conversation topic. Further, an AC could “bookmark” a part of the conversation (e.g., using a short buffered audio recording, or an extracted phrase) to remind other speakers of the original context of a delayed response.

One-speaker constraint: ACs with utterance-level voice output made clear to all speakers that they held the floor and did not experience any missed opportunities caused by erasing a message they wanted to share. Future work may consider other techniques of augmenting ACs' communication with signals that provide insight about ACs' intent to contribute. Prior work that visually displays an AC's status to other speakers (*e.g.*, typing, idle, listening) [46, 146] may help decrease the impact of the one-speaker constraint on the AC by encouraging other speakers to take into account the AC's status. Future work may also consider discrete augmentations that cue partners without distracting their visual attention from the AC, such as wearables or peripheral signals in the environment that are accessible in space and beyond the screen.

3.6.2 Designing for Relationship, Device and Task Constraints

Relationship constraint: Similar to Blackstone's "Circles of Communication", our study suggests that ACs communicate differently with different partners [15]. For instance, ACs used nonverbal communication more frequently with their close conversational partner (CCP) than they did with the third party (TP). As CCPs often translated AC nonverbal communication to the third party, future devices could also translate or amplify AC nonverbal communication to other unfamiliar third parties to expand initial communication modalities. We observe beneficial interdependence [9] between the AC-CCP pair, and acknowledging the unique importance of the contributions of each communicative partner can inform new interactions. Considering interdependence, we could extend AAC groupware [46] to consider tools designed specifically for users with different relationships beyond the CCP (*e.g.*, a privacy preserving application for new partners), or for communication groups with many participants.

Device constraint: Device output type (*e.g.*, utterance-level vs phrase-level) impacted ACs' ability to hold the floor once they started speaking. Device input type (*e.g.*, scanning vs direct selection) contributed to the effort it takes for an AC to speak, and prior work suggests that ACs consider whether speaking is "worth the effort" before participating [77]. In the future we should continue to improve device throughput, but also educate non-AC users on how better support AC participation in conversation.

Task constraint: When the CCP knew the information needed to complete a task, the CCP supported the AC through informed participation. When the CCP did not know the information needed for the task, the CCP instead guessed and asked questions to help the pair achieve the goal. Future work could consider how to identify a conversational goal and relevant information from context to better support ACs' expression of agency. For instance, AAC systems could also try to detect and suggest possible general goals of the conversation that could help inform conversational regulation strategies that promote accomplishing this goal.

3.7 Summary

The conversational agency in AAC framework adds to the body of AAC work in HCI by considering the role of close conversation partners under contextual constraints that, when loosened, create opportunities for augmented communicators to exercise agency. Identifying conversational, task, device and relationship constraints can generate new augmentations that extend beyond the device and to enable different conversational dynamics that favor the expression of agency of augmented communicators.

Part II

Agency through Motion

Chapter 4

Co-Designing forms of Motion-based AAC through Robotic Sidekicks

AAC devices enable speech-based communication. However, AAC devices do not support nonverbal communication, which allows people to take turns, regulate conversation dynamics, and express intentions. Nonverbal communication requires motion, which is often challenging for AAC users to produce due to motor constraints. In this work, we explore how socially assistive robots, framed as “sidekicks,” might provide augmented communicators (ACs) with a nonverbal channel of communication to support their conversational goals. This chapter describes two separate studies to explore motion as a new communication modality for AAC.

In the first study we developed and conducted an accessible co-design workshop that involved two ACs, their caregivers, and three motion experts. We identified goals for conversational support, co-designed prototypes depicting possible sidekick forms, and enacted different sidekick motions and behaviors to achieve speakers’ goals. We contribute guidelines for designing sidekicks that support ACs according to three key parameters: attention, precision, and timing. We show how these parameters manifest in appearance and behavior and how they can guide future designs for augmented nonverbal communication.

We then carried out a second study (Chapter 5) in which we build from this prior work to dig deeper into a particular case study on motion-based AAC by co-designing a physical expressive object to support an AC during group conversations. We found that our physical expressive object could support communication with unfamiliar partners and we present our process and resulting lessons on the designed object itself and the co-design process.

4.1 Overview

In Human-Robot Interaction (HRI), social robots have used motion to support people in a variety of related communication contexts through nonverbal behavior. For example, researchers have used social robot gaze to facilitate turn-taking [2, 106], and robot body gestures to mediate interpersonal conflict [68, 150]. In this work, we leverage robots’ advantages of embodiment and motion for nonverbal communication in the domain of AAC, to explore how social robots could be designed to support ACs’ nonverbal communication. We use the metaphor of “assistive sidekick” to explore robot designs that *support* ACs, rather than function as independent agents.

To investigate how social robots could be used in augmented conversations, we conducted an accessible, multi-phase co-design workshop, with the goal of designing nonverbal behaviors that are rich and understandable enough to support augmented speakers’ conversational goals. Participants in the workshop included ACs and their caregivers, also referred to as close conversation partners, in order to set the co-design process around their experiences. It also included puppeteers, who contributed their expertise in expressive motion to explore new motion-based behaviors for



Figure 4.1: We explore the metaphor of an “assistive sidekick” to design robots that assist augmented communicators by playing a supporting role through nonverbal behavior.

assistive robots.

Our work makes three primary contributions to socially interactive robots for accessibility and HRI design research. First, we offer an example of a co-design workshop adapted to include the participation of ACs with motor disabilities, who are often excluded from design processes. Second, based on the workshop process and outcomes, we contribute guidelines for designing social robots that support ACs through expressive nonverbal behavior. We describe these guidelines according to three key aspects: attention, precision, and timing in conversation. Third, we identify a list of potential conversational goals important to ACs that “assistive sidekicks” could support, and a set of motions and behaviors to achieve them.

4.2 Method

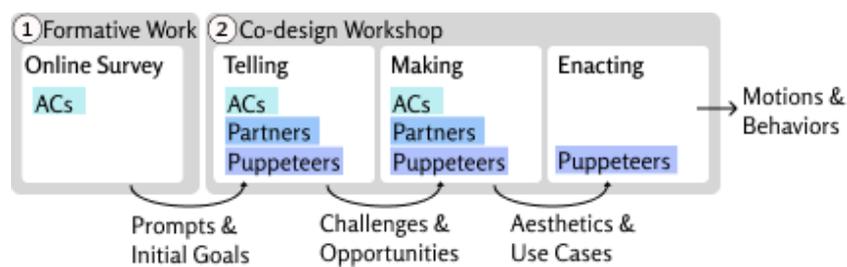


Figure 4.2: Our two step process included formative work and a three-phase co-design workshop. Participants included augmented communicators (ACs), their close conversational partners, and professional puppeteers.

We conducted our research in two steps (Figure 4.2) to answer the following research questions: (1) What are ACs’ nonverbal needs in conversation?; (2) How, if at all, can a sidekick support ACs in conversation? The first step consisted of gathering information about current ACs’ needs in conversation via an online survey distributed to a group of ACs. The answers collected in the online survey served to identify interaction challenges a sidekick could help with. The second step of our process consisted of an in-person co-design workshop split into three phases: *Telling*, *Making*, and *Enacting* [21]. *Telling* activities focused on scaffolding the sharing of stories of current lived experiences, reflecting on what mattered to ACs in communication, and identifying potential challenges to address. *Making* consisted of externalizing ideas and embodying

solutions in the form of physical artifacts to define possible sidekick forms and aesthetics. Lastly, puppeteers *enacted* and explored sidekick motions through behavior improvisations.

4.2.1 Step 1: Online Survey

Four augmented communicators (aged 33 to 54) with motor disabilities responded to our survey. Respondents use switches, eye gaze, touch screens, and keyboards to access their AAC devices, and all have more than 5 years of experience using AAC. The survey included a combination of Likert scale and open-ended questions asking participants to rate how often they experienced specific situations in conversation and to describe them further.

Three of four participants indicated that they participate in conversations less than they desire to. Participants expressed not being able to communicate fast enough to keep up with topic transitions, a finding that confirms prior work [125, 157]. Participants reported feeling pressure to respond quickly in an ongoing conversation and often asking their caregivers to explain things on their behalf.

Participants also shared that people who are close to them occasionally help facilitate their communication by telling others to be patient, to let them participate, and to address them directly (“They tell [people] to talk to me”, “[I like that] my partner looks [at] other people to make sure they are communicating with me”), “It’s nice to have someone who knows I am typing and will tell everyone.” Sometimes close conversation partners use hand gestures to make another person wait until the AC finishes their thought (“They look at me the whole time with their pointer finger extended out, as if she is saying [to the other person], ‘Hold On’.”)

Based on the online survey and along with prior work [157], we identified that it would be worthwhile to explore whether an embodied technology could play a supportive role similar to a knowledgeable conversation partner and communicate that (1) an AC would like to participate, (2) that they should be addressed directly, or (3) that they are composing a message.

We consulted with a speech-language pathologist (SLP) who confirmed the identified challenges as known problems, and suggested other critical tasks partners help with, such as telling others that an AC is having trouble with their device. We used these findings to create an initial list of goals a sidekick could support (Table 4.1).

We framed the envisioned technology as “assistive sidekicks,” physical expressive objects that play a supporting role to the AC, ensuring an AC’s conversational agency without replacing them or their knowledgeable partner. We then designed a workshop to explore this envisioned technology.

4.2.2 Step 2: Co-design Workshop

We designed an accessible workshop that included three phases: (1) *telling* to uncover important design goals and parameters, (2) *making* to build potential sidekick forms, and (3) *enacting* to design sidekick motions and behaviors. The workshop was 6 hours long.

While co-design workshops provide benefits for aggregating perspectives and expertise in design [128], some design activities can be inaccessible to ACs who have speech and motor disabilities. Co-design activities can heavily rely on the ability of participants to verbally express their ideas in a timely manner and to physically produce sketches or prototypes. We therefore designed a co-design workshop with accessibility in mind, to ensure that ACs’ ideas would be included in the process. Our goal was to accommodate the range of speaking rates, styles and preferences of our participants. First, we established everyone as a collaborator and used making as a form of ideation [87]. Next, we developed a series of *Access Commitments*, following prior work and in collaboration with an SLP [11, 23, 87]. Our *Access Commitments* included: (1) minimizing participant fatigue by planning three breaks between each phase and allowing asynchronous discussion after the workshop, (2) sending question prompts in advance, (3) establishing preferred modes of

communication, (4) allowing enough time for everyone to communicate during the workshop, and (5) making visuals and materials easy to reference and within reach for ACs.

Workshop Participants Workshop participants (for whom we use pseudonyms) included two augmented communicators (ACs), their close conversation partners, and professional puppeteers. Tammy and Matt are both AAC device and wheelchair users (aged 38 and 23 respectively). Tammy participated in the workshop with her dad, and Matt participated with his mother and his aide. Both have cerebral palsy (CP) and more than 18 years of experience using AAC devices. Tammy uses a joystick to point to words on her device and pauses over her target to select. Matt's device constantly scans through all options on the display (e.g., word, letter, shortcut) until he presses his head switch to select a desired option. Matt needs more time to use his device than Tammy. Matt and Tammy also use some nonverbal behaviors to communicate. Tammy uses her hands to make pointing gestures and signs she and her family have agreed on. Matt uses arm restraints because of his CP, but makes use of facial expressions and eye blinks to support communication.

We additionally invited three artists with prior experience in puppetry, art and puppet design. Puppeteers understand nuances of movement and use motion to elicit specific interactions and responses. Puppeteers have previously been involved as movement experts in the design and research of nonverbal robots [143, 98].

Telling: Uncovering Challenges & Goals The first phase of the workshop aimed to discover ACs' needs and challenges in conversation that would benefit from nonverbal support. We carried out brainstorming and story sharing activities to uncover opportunities and tensions related to communication, and to hear prior successful and/or challenging communication experiences. These served as a first step toward identifying potential goals for a sidekick.

In the first activity (brainstorming), we asked participants to write their answers to the question: "What does communicating mean to me?" and to choose a metaphor they related to communication. We used a selection of "Metaphor Cards" [95] and asked participants to choose one to share with the group. Metaphors can allow participants to think about abstract concepts in new ways, and to create novel discussions on a topic [95]. Metaphor Cards were numbered to allow ACs to easily reference them without pointing. This activity aimed to uncover some of the current tensions, hopes, and challenges in ACs' communication.

The second activity (story sharing) aimed to explore and discuss AC needs with the group. We asked ACs to share stories about their face-to-face conversations which highlighted challenging, successful or humorous moments. This raised discussions about potential conversational goals a sidekick could assist with (Table 4.1).

Making: Aesthetics & Use Cases Drawing from the stories shared, participants prototyped assistive sidekicks on two AC-led teams. The goal was to explore how a sidekick might address the needs and values identified during the *telling* phase, and to think-through-making by rapidly prototyping ideas in a group [109]. We asked participants to select one challenge that resonated with them from the discussions, and to use craft materials to create an assistive sidekick that might alleviate the identified challenge. We provided materials to create basic structures, and decorative material such as fabric, feathers, and googly eyes. We divided participants into two teams, each led by an AC and including their caregiver and one or two puppeteer(s). After prototyping, teams demonstrated their sidekicks, described the challenge it would alleviate and the behaviors it could leverage to do so.

Enacting: Developing Motions & Behaviors The third part of the workshop aimed to design sidekick motions and behaviors that could meet specific goals. Puppeteers enacted situations and behaviors using improvisation as a form of bodystorming [134] to explore how previously identified AC needs would play out through motion and behavior. Similar improvisation with

domain experts has been previously used in HRI research to explore motion for robotic prototypes [142, 143, 174]. We guided puppeteers through a behavior improvisation session that built upon what they had learned from ACs and their partners to generate concrete behaviors for supportive nonverbal sidekicks.

To minimize AC participants' fatigue we decided not to include ACs in this last phase of the workshop. Only puppeteers participated in enacting specific sidekick motions. Researchers sent videos of the generated behaviors to ACs for their feedback post-workshop.

The researchers selected four identified conversational goals for the puppeteers to explore, based on their prominence in discussion and their diversity. The selected goals appear in bold in Table 4.1. One goal at a time, we instructed puppeteers to generate as many behaviors as they could to achieve it. For example, puppeteers were instructed to show many ways in which a sidekick might communicate to the conversation partner that the AC would like to add something to the conversation. After exploring possible appropriate motion gestures for each goal, we held a brief discussion with puppeteers to talk about what worked, what did not, and what they noticed in the process of improvising motion to support nonverbal supportive goals for sidekicks.

Data Collection and Analysis All research procedures were approved by our university's institutional review board. All AC participants were recruited via AAC interest group email lists. Puppeteers were recruited through art school referrals. Workshop participants were compensated with 20 USD per hour. Survey participants were compensated with a 30 USD gift card.

We captured video, audio and photographs during the workshop. Three paper authors additionally took observational notes including participant quotes and paraphrasing of events with timestamps. We used Affinity Diagramming to perform thematic analysis [103] on the notes post-workshop. To analyze the behavior improvisations, we focused on identifying emergent themes from the puppeteer-generated motion, also using thematic coding. To evaluate how accessible our workshop was, we examined ACs' participation by evaluating the turns they took as measured in prior work [157]. We also directly asked ACs about their workshop experience and inquired suggestions for improvement.

4.3 Workshop Findings

The workshop resulted in several findings: (1) Design parameters to guide the design of nonverbal communication sidekicks: *attention*, *precision*, and *timing* in communication; (2) a refined set of AC conversational goals that a sidekick might support (Table 4.1); (3) a set of physical prototypes to represent potential sidekick forms (Figure 4.3); and (4) a set of motions and behaviors that the sidekick could use (Table 4.2). These four findings tackle four layers of this design space, one building on top of the other. Together, they form a coherent picture of how one might design robotic sidekicks to support ACs' nonverbal behavior.

4.3.1 Attention, Precision, Timing: Parameters for Nonverbal Communication

We begin by defining the design parameters that manifested throughout our workshop. These parameters do not encapsulate all challenges ACs face, but can serve as design guidelines to be explicitly considered when designing sidekicks for AAC.

The *attention* parameter is concerned with ACs' need to be able to draw attention to themselves, to divert attention towards something other than themselves, and to maintain their partner's attention during an interaction. Our ACs reported struggling with calling attention to themselves or fully participating in conversation because of motor and speech impairments. They reported

that it is difficult to produce a loud sound, raise a hand to participate, or point to something—an attention challenge that echoes prior work [147]. Co-designers expressed excitement about the sidekick’s potential to use motion to call for attention during communication. Participants shared how they could use the sidekick to signal for attention in crowded and loud places by waiving at others or giving others a funny look. Caregivers and ACs discussed that being able to divert attention towards something other than the AC was also important. One of the caregivers commented: “[when there is a long silence] the person over here is like, can I come over to the other side and look?” Her comment sparked a discussion on a potential use of a sidekick’s nonverbal behavior to divert a conversation partners’ attention to ACs’ screens. However, ACs also expressed hesitation about drawing too much attention, noting that sidekicks should balance bringing attention to the ACs while at the same time not being “too distracting.”

The *precision* parameter describes two conversational functions: (1) the need to add nuance to a spoken message to more precisely match an AC’s intention and (2) the need to add clarity to the AC’s existing nonverbal communication. Participants suggested that a sidekick could help convey the AC’s tone and mood when needed, like acting in a playful way to convey lightheartedness. Co-designers also discussed how ACs’ nonverbal communication can be faster than speaking with the AAC device, but also more ambiguous. Tammy explained: “One day my mom taught me to point to my throat when I was thirsty. I was thirsty so I tried my new sign on dad and he got a funny look on his face. He rubbed [ointment] all over my neck, it was a while before I got a drink.” Thus, sidekick nonverbal communication should be precise, as an ambiguous message can come at a great cost and be difficult to repair.

The *timing* parameter describes the level at which communication can be conveyed in the exact moment that is needed. Being able to say things in time is a challenge for ACs, given that operating an AAC device is slower than producing verbal speech. ACs expressed a need to be able to control timing, such as when to take a turn during a conversation, or when to ask for additional time. Matt’s mother reported: “Matt has so much stuff pre-programmed in there [AAC device], but the problem is that he can’t get to it very quickly so it doesn’t work even though it is already there... he can’t get there fast enough”. To address Matt’s communication timing issue, Matt’s family created a light that is placed on his wheelchair, close to his head. Matt turns the light on with a head switch when he is ready to speak, for example, in his college classes. Matt’s light is one example that allows him more control over the timing of his communication by non-verbally alerting other speakers. In addition, the timing of when a sidekick displays a nonverbal behavior, either before the AC speaks or during their turn as to accompany their message, was further explored during puppeteer improvisations.

In the next section we describe how these three design parameters were manifested in each stage of the sidekick co-design process.

4.3.2 Conversational Goals for Sidekicks

The first two workshop phases (*Telling* and *Making*) allowed us to construct and refine a list of possible conversational goals for a robotic sidekick (Table 4.1). These goals address conversational and social challenges experienced by ACs and partners when having a conversation involving AAC.

Stories shared by close conversation partners elaborated on ACs’ communication challenges. For example, Matt’s mom explained how even though she is familiar with AAC devices, some words still sound harsh: “I was with an AAC user and we were having a talk. We were doing something, and she said “I have got to go now. Goodbye.” I found myself feeling hurt when she said it. There is an aspect of the computer that made it abrupt.” This experience is related to the parameter of precision. The verbal communication from the AAC imprecisely conveyed the AC’s feelings, so Matt’s mom did not know if the AC meant to say goodbye in a warm or cold way. An accompanying nonverbal signal might have added precision to the verbal utterance by augmenting

Table 4.1: Goals along with their corresponding numbers and the stage at which they were introduced: original goal from formative work (O), goal mentioned during the *telling* phase (T), goal mentioned in *making* (M). Goals 1, 5, 8, and 10 were selected to be further explored in the *enacting* phase.

#	Stage	Goal
1	O	Show others the AC wants to participate
2	O	Encourage addressing the AC directly
3	O	Show others there is an AAC device problem
4	O	Show others the AC needs more time
5	T	Fill the silence gap while the AC types
6	T	Show others the AC is ready to communicate
7	T	Tell others to remember what they're talking about
8	M	Soften an AAC device message
9	T	Show others that the AC is typing
10	T	Invite others to look at the AC's screen
11	T/M	Show others the AC disagrees
12	M	Add humor to an AAC device message
13	M	Convey emotion quickly

it expressively.

Tammy's father described another challenge: the difficulty of remembering a conversational topic due to the time it takes ACs to craft a message. He explained that Tammy's subtle head movement is a useful cue to know when to remember the topic of conversation: "*I tell others to watch when Tammy looks down, that means she is typing, and we need to remember this is what we are currently talking about.*" Tammy's natural movement of lowering her head serves as a timely marker. Nonetheless, it requires an attentive partner who is familiar with Tammy's nonverbal behavior to be able to recognize this gesture.

These examples, along with other stories told by ACs and their caregivers highlighted many opportunities for nonverbal behaviors to support ACs' communication and agency (Table 4.1). We used the identified needs list as a "design space guide" for the next exploration stage of our work. Furthermore, we hope some of the needs identified here will guide future work on the design of nonverbal robotic sidekicks for AAC.

4.3.3 Sidekick Aesthetics & Use Cases

The prototyping session during the workshop's *Making* phase turned the focus to an assistive sidekick's appearance and function, while setting aside concerns of technical feasibility. In this activity, form followed function as the teams first decided on a need from the identified needs list before building it from craft materials. The prototype designed by Tammy's team (Figure 4.3, left) had anthropomorphic features such as eyes, eye brows, and hair, while Matt's (Figure 4.3, right) resembled more of a "protest sign" object with a message that would move and appear as needed.

Adding Emotion Through Facial Gestures. Tammy's team decided to address the need to convey emotions quickly in times when it is difficult to use an AAC device, like in noisy and crowded places. This design goal touched primarily on the *attention* and *precision* parameters. The sidekick was intended to capture and express Tammy's sense of humor in communication (add precision to her messages) and to be visible from different angles (to grab attention). The team prototyped an expressive face, considering gaze and eyebrows as legible ways to quickly convey emotion. Tammy's dad mentioned how the eyes could be shortcuts for expressions (e.g., rolling eyes or "bored" droopy eyes). Tammy playfully said she wanted to use her sidekick to express "you are pissing me off". Tammy's team spent time customizing the prototype's form to represent Tammy, for example matching its eye color to Tammy's eyes.



Figure 4.3: Prototypes by Tammy's (left) and Matt's teams (right). Tammy's prototype featured a large set of eyes, customized to express sentiment. Matt's prototype read "one moment please" to communicate a desire to contribute.

Signaling Disagreement Through Object Appearance. Matt was primarily interested in having a way to express that he disagreed and needed a moment to share his point of view. To accomplish this goal, Matt's team designed an object that resembled a protest sign, carefully designed to show respectful disagreement. Matt's goal touched upon all three of the design parameters. In his optimal scenario, the sidekick would convey that Matt disagrees in a precise way that does not seem disrespectful. It would draw attention to let others know he wants to contribute. Finally, it would use the right timing and give Matt time to prepare a response. In order to convey respectful disagreement, the team used a magenta-glitter background and furry fabric, which served as a lighthearted visual signal that aims to soften the message. The use of a different language ("Un momento por favor" in Spanish) also increases playfulness. To achieve drawing attention towards Matt, the team used bold colors, and a shape of a pointing finger with its nail painted black. If necessary, the prototype could rest close to Matt's face, pointing directly at him for additional attention and lowering itself to a hidden position when not in use.

4.3.4 Sidekick Behaviors and Motions

In the final workshop activity (*Enacting*), puppeteers improvised different behaviors to meet four conversational goals identified in prior activities (Table 4.1), with the newly generated prototypes. The selected goals were: (1) filling the silence gap while an AC types and others wait, (2) showing the AC’s intent to participate, (3) softening a message spoken by the AC, and (4) inviting others to look at the AC’s screen. We explored these few goals in-depth as an initial examination of this complex design space, prioritizing a deeper exploration of possible behaviors per goal. In future work we would like to examine additional needs and goals. Puppeteers improvised using the prototypes created by Tammy and Matt, but sometimes also used their own bodies or other props. In this section we describe each explored goal, the motions used to achieve it, and the motion patterns that emerged.

Table 4.2: Puppeteers generated 10 behaviors for each selected goal using sidekick prototypes of their choice. We highlight three example behaviors per goal.

Goal	Behavior Title	Behavior Description
Fill silence gap	Tapper	Taps fingers to show typing
	Dancer	Dances and sings
	Boomerang	Moves cyclically out from the AC and back to original position
Let AC participate	Up Up Up	Moves up for attention, then opens mouth when partners look at the AC
	Attention Sign	Springs up sign from flat to upright
	Outward Reach	Moves forward for attention, gazes at partners who are still speaking
Soften a message	Gentle "No"	Spins to soften "no"
	It's Time	Looks at watch patiently to soften "I have to go"
	Bye-bye Wave	Waves to soften "I have to go"
Invite partner to look at screen	Imitate Me	Demonstrates looking at screen
	Inviting Arm	Gestures with an arm to invite partner to look at screen
	Swirly Look	Spins then looks at screen

The puppeteer improvisations shared a common structure, composed of two “stages” presented in sequence. First, they performed a motion to call for attention. Such motions included making the sidekick take up more space by moving upwards or outwards, spinning or moving quickly, or appearing from a hidden place. After attention was captured, puppeteers tried to convey the main message or purpose of the gesture by either mirroring an action that the AC would be doing (e.g., mimicking typing on a keyboard) or by demonstrating something that they would like the conversation partner to do (e.g., look at the AAC device screen to invite the partner to do the same).

The motions improvised by puppeteers also reflected our identified design parameters. First, *attention* was a key function for motion that preceded conveying a message. The sidekick’s movement and physical presence supported drawing attention to the AC. The second part of the behavior, either mirroring the AC or suggesting what the conversation partner should do, served to

clarify the message and make it more *precise*. These behaviors also considered the right *timing* by presenting themselves either in response to an AC's action (e.g., while the AC typed) or in parallel, as the AC spoke their message. Puppeteers generated 10 different behaviors for each goal, a total of 40 different behaviors. We include three example behaviors for each goal in Table 4.2.

In order to **fill the silence gap** that occurs when an AC types and a partner needs to wait for a response, puppeteers made sidekicks move rhythmically in circles (like a “loading” icon) and mimicked typing on a keyboard to show progress. Other explored behaviors aimed to fill the silence through entertainment, such as having the sidekick give the partner a back massage or dance and sing. The *precision* parameter was prioritized for this goal—puppeteers agreed that more familiar gestures like typing or “loading” were easier to understand.

To **show that the AC wants to participate** the puppeteers had to first find an effective way of interrupting an ongoing group conversation (played by other puppeteers). *Timing* was revealed as the most complex for this goal—it was difficult for puppeteers to identify a good time to interrupt a busy group. Some motions that explored ways to call a group's attention included moving the sidekick upwards so that everyone could see it, making the sidekick appear in a sudden manner, or moving outwards to a central focal point of the group. Once the prototype caught the group's attention, the puppeteers attempted to communicate the intent by pointing at the AC, gently poking the partner, or by using gaze to look at the AC to indicate with precision that the AC wants to participate.

The third goal explored how a sidekick could **soften a message**. While nonverbal behavior allows people to set the tone of a verbal message, monotonic AAC device speech can make this challenging. Puppeteers explored adding emotion to a message through slow motion, head tilts, and nodding. To soften the message even more, puppeteers used signals of empathy, like sending a kiss or patting the partner's back. Expressing emotion allowed clearer communication of intention. Being able to soften a verbal message is primarily concerned with *precision*—the motion is intended to communicate the AC's intention more accurately. *Timing* was connected to the level of *precision*—varying when and at what rate motions are presented could alter the perceived level of “softness” or “harshness.”

The last explored goal intended to **invite a communication partner to look at the AAC device's screen**. Similarly to other goals, puppeteers achieved this in two steps: catching the partner's attention and then making a gesture towards the screen. Different motions successfully captured the partner's attention, like spinning the sidekick or making it appear from behind the screen. Once the sidekick had the partner's attention, it could indicate for them to come and look at the screen. This was achieved by pointing the sidekick towards the screen, using gaze, or using the sidekick's head to point. Achieving precision was challenging for this goal, as it did not necessarily accompany a verbal message by the AC, but rather communicated its own message.

We shared videos of the puppeteers' improvised behaviors with our two AC participants after the workshop. Only Matt provided additional input. Behaviors that were “straightforward,” as described by Matt, were positively rated such as the *Tapper*, *Attention Sign*, *Inviting Arm*, and *It's Time*. Other behaviors were labeled “too distracting,” like the *Dancer*, and were therefore disliked.

4.4 Workshop Accessibility Reflections

As an additional contribution, we report on the successes and challenges of our five access commitments (Section 4.2.2).

Our **first commitment** aimed to minimize participant fatigue by planning breaks between activities and allowing asynchronous discussion post-workshop. This had both positive effects and challenges. Breaks were used and appreciated by all. Despite our attempts, there was less engagement post-workshop, with only Matt responding with feedback about puppeteer-generated behaviors. Matt uses his computer every day for work which was not the case for Tammy. We

recommend future work to acknowledge differences in computer usage, and to tailor post-activity follow-up strategies for each participant’s usage context.

The **second commitment** was to send question prompts that will be used in the workshop in advance. Both AC participants shared stories that were pre-recorded on their device, suggesting they made use of the prompts that were sent in advance.

Our **third commitment** allowed time to establish preferred modes of communication. We found that asking participants to share communication preferences and strategies at the beginning of the workshop prevented possible misunderstandings. Matt’s mom explained that Matt uses blinks to communicate “yes”, “no” and “*I don’t know.*” Matt’s team then used these cues when working on the prototype. Matt’s mom also explained that Matt gives 1-2 word answers and asks others to elaborate on his behalf. Matt used this strategy in the workshop to ask his mother to elaborate for him. Similarly, knowing that Tammy puts her head down to indicate that she wants to type, as her father explained, helped a puppeteer unfamiliar with AAC realize that Tammy was busy typing and that she should wait before asking her next question.

Our **fourth commitment** was to allow everyone enough time to communicate during the workshop. We took turns in an established order, which supported ACs and gave them more time to communicate. For each activity, all participants took at least one turn. One strategy that emerged was to first ask Matt and Tammy the question, but then let other participants take their turns as Matt and Tammy prepared an answer. This encouraged longer responses from AC participants. Another emergent strategy was for each team to confirm all design decisions with the ACs. Matt’s team asked Matt a series of questions for feedback. For instance, after Matt said “I want to use what we were talking about,” the team asked follow up questions and, with Matt’s blinking to answer yes or no, narrowed the topic down to find out what Matt meant. Tammy created a sketch while her team consulted with her on key decisions.

Our **fifth commitment** was to make all materials easy to reference and access. Labelling visual materials with numbers and annotating discussions on a board enabled ACs to use verbal shortcuts to refer to specific concepts. For example, Matt used a number to indicate his chosen metaphor instead of having to type out the metaphor name. To support physical access, we used adjustable-height tables that allowed ACs to get close to prototyping materials. However, we lacked other useful adaptive tools (e.g., head-mounted tools) which may have eased prototyping. We recommend labelling all visual materials and asking participants about desired access tools in the process of designing a workshop.

4.5 Discussion & Future Work

We have identified opportunities in which socially assistive robots could augment AAC-based communication using motion. Robotic sidekicks are not the only solution, but provide an interface to explore motion-based nonverbal communication, which is missing from current speech-based AAC technology. In this section, we reflect on how social robots’ motions could improve precision, timing, and attention in communication, and identify further research questions that can guide the design of functional robotic sidekicks.

4.5.1 Precision through Embodied Motion

The sidekick’s motions should be immediately recognizable and interpretable so that people who may not be familiar with AAC or an AC’s communication style can still understand the message [15]. Most behaviors produced by puppeteers in our workshop leveraged common, established nonverbal gestures to convey a particular concept. These included gaze and eye motion, head orientation, pointing, and iconic gestures such as looking at a watch to communicate “time.” Leveraging such “universal symbols” was helpful for quick communication that can speak to a range of audiences, as puppeteers explained when they reflected on common gestures. That said,

gestures are likely to be culture-specific, and would need to be re-examined depending on the user and audience. Precision through embodied motion is only one approach; other modalities, such as expressive voices [45], may add further precision.

We learned that specific aesthetics can also support conveying a precise message. Tammy’s team created an anthropomorphic sidekick with highly visible facial features that could express familiar facial expressions. Matt’s sidekick, shaped as a sign, suggested a strong and direct presence, though it also used its aesthetics to convey friendliness. Future work could examine visual aesthetics of sidekicks and AAC technology as an additional modality for increasing the precision of a message. The final designs may have been shaped by the materials available at the workshop as well as personal preferences, so additional explorations with other materials and AC participants could reveal additional sidekick forms.

4.5.2 Timing of Sidekick Behaviors

The timing of our proposed sidekick behaviors showed both *consistency* and *flexibility*. Behaviors followed a consistent two-stage sequence, starting with a motion to call for attention, and then taking some communicative action. However, behavior timing was also flexible depending on the context of the interaction. Some behaviors occurred before the AC verbally communicated, in order to fill a silence gap or to ask the partner to view the AC’s screen. Other behaviors overlapped with AC speech, as when the nonverbal behavior was intended to soften a message or increase a message’s precision. Finally, some behaviors occurred at short intervals in communication, for example to let the group know that an AC is ready to speak. Identifying the right time to perform a behavior requires awareness of context and conversational dynamics.

In our workshop, we sidestepped this timing issue by implicitly assuming the sidekick would be activated at the right time. However, leaving the AC to activate a sidekick’s behaviors at the right time is likely to increase their cognitive and physical burden. On the other hand, having a sidekick autonomously recognize interaction context to activate its own behaviors is itself a major research challenge. Future work could examine tradeoffs between user-activated and autonomously activated sidekick behaviors.

4.5.3 Modalities for Gaining Attention

Throughout the workshop, it became clear that one of the most important abilities for a sidekick is to gain attention from conversation partners. This was evident not only in the behaviors explored, but also in the physical prototypes created by the co-design teams. Both teams used embodied properties to gain attention in their prototypes: Tammy’s team used eyes, a feature that is known to draw attention, and Matt’s team used components with high visibility—glittery pink paper and furry fabric. Puppeteers also varied motion speed and gesture size to call for more or less attention as needed, an approach similar to prior work [1].

Matt’s current use of his wheelchair light to signify his desire to speak highlights an opportunity for sidekicks to gain attention through more ambient, non-anthropomorphic modalities. This might be especially useful in situations where anthropomorphic motions are inappropriate or impractical. Future work may investigate the benefits of different attention-getting modalities for expressive sidekicks in challenging contexts, like noisy environments.

4.5.4 Building Motion-Expressive Sidekicks

Translating the designs from this workshop into fully functional robots will require additional work and design. For example, the sidekick should have accessible controls (e.g., eye gaze, infrared remote control, or switches) so an AC can trigger different behaviors. It is important that these interfaces do not impose additional burden on ACs. Future work could explore what type of

control is most desirable to ACs, potentially extending control to include the partners (as with groupware [46]) or robot autonomy (as with shared autonomy[73]). If the sidekick has any amount of autonomy, it will need to use techniques from sensing and automation to select context-aware behaviors. This is still an open research challenge in social HRI. The sidekick might also use machine learning to personalize its behavior to a single individual over time.

4.6 Summary

Our findings offer design opportunities to enhance augmented communicators' (ACs') face-to-face conversations by using socially assistive robots as a form of motion-based AAC. In a co-design workshop with ACs, their close conversation partners and puppeteers, we explored the metaphor of a "sidekick"—a physical device that supports ACs in conversation through movement. We provide evidence of the usefulness of motion to bring attention to ACs, convey precise messages, and enable ACs to participate in conversation at the right time. We present key needs that side-kicks could successfully address, as well as accessible co-design commitments that amplify ACs' participation in co-design.

Chapter 5

Aided Nonverbal Communication through Physical Expressive Objects

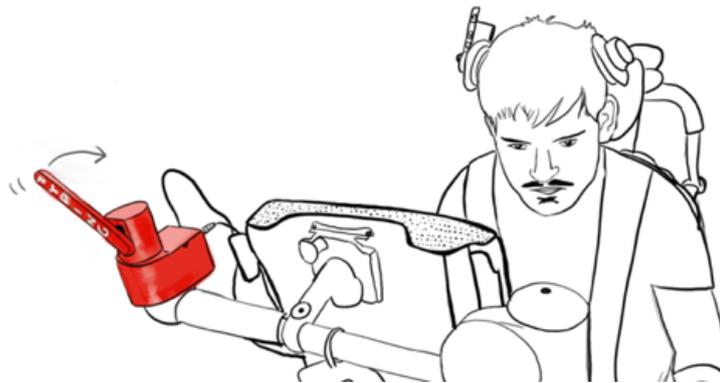


Figure 5.1: A design team including an augmented communicator, his close conversation partners and HCI researchers worked together on designing a physical expressive object to support AAC-based communication.

5.1 Overview

In this study we engaged in participatory design [128] to co-design a bespoke technology with an augmented communicator and his close conversation partners. We carried out research through design, in which we engaged in designing as a research activity to obtain design knowledge in the process [172], positioning design activities as data collection and analysis opportunities [113]. Our activities included information gathering through interviews, surveys, prototyping sessions, and diary entries over a period of 12 months. We documented our decision-making process that occurred over several in-person and remote sessions. We captured every conversation, meeting, and activity with notes, daily reflections, artifacts, or video and audio recordings when permitted among all involved co-designers. We collected all online exchanges including pictures, emails, and discussions we had on remote collaboration tools. All audio recordings were transcribed and artifacts were analyzed to find common themes.

5.2 Design Team Participants

This work involved a design team including Mark, an expert augmented communicator who is also a co-author of this work, Mark's close conversation partners, and university HCI researchers. We describe our participants Mark and his family members, T and D.

Mark has used AAC for over 19 years and has long advocated for AAC users by working at a world leader AAC device company, serving as a student government senator, and advocating in a variety of disability rights campaigns. Mark has cerebral palsy and is a wheelchair user. He uses indirect selection to control his AAC device. That is, Mark's AAC device scans through each option on the device (e.g., word, letter, shortcut) until Mark presses his head switch to select a target. In addition to his device, Mark uses facial expressions and eye blinks to communicate.

T and D facilitated our remote design meetings, sharing about their experiences communicating with Mark, and facilitating the exchange of materials needed to carry out this work. University HCI researchers. HCI researchers with backgrounds in accessibility, fabrication, and robotics coordinated the design activities.

5.3 Design Process

This section describes our design process, detailing the design decisions made and highlighting design values we needed to consider when designing a motion-based form of AAC. We describe: (1) our discovery of design opportunities for nonverbal expressive sidekicks, (2) the definition stage during which we set out to discuss design priorities for a future functional sidekick, (3) the development of different ideas and prototypes, (4) a delivery stage in which we pilot tested the device in context and refined our prototype, and (5) the evaluation stage during which we tested the prototype for two months, with three weeks of diary study collection.

5.3.1 Discovery: Opportunities for Expressive Objects in AAC

We invited Mark to co-design a physical expressive sidekick for his use. Mark had been a participant in a past study and had expressed interest in exploring a solution that would allow him to better capture his communication partner's attention when needed and without interrupting the ongoing conversation. Mark also noted that he would mainly want to use his sidekick during group conversation, in a classroom setting or student government meeting. Some challenges he experiences with group conversations involve turn-taking such as knowing when to interrupt a group or being able to tell others he is working on a message before conversation partners assume he has nothing else to contribute and move on to the next topic. The time it takes Mark to compose a message can vary between a couple of seconds to up to minutes and therefore he often asks for more time to compose a message to remind partners to wait.

At the time we started this work, Mark was completing his last semester of college and was attending student government meetings for which he acted as vice-president. He invited us to observe one of his meetings so we could get a better sense of how his current group conversations developed. Unfortunately, the COVID-19 pandemic commenced during this time and the student government meetings were canceled. Mark shared via email a little more about the in-person meetings:

“The conversation is fast paced, and sometimes it's difficult for me to contribute. However, the other senators are tuned into how my communication system works. As vice-president, I'm responsible for keeping time so I have the following preprogrammed phrases: 'I'm sorry to interrupt, but we're running out of time; Let's get back on track. We can always discuss this at a later time if necessary.' 'I would like to add something to the conversation. It will take me a few minutes to write it. Would you call on me in a few minutes?’”

Upon further discussion with Mark and his family, we learned that he uses a variety of strategies to let others know he wants to compose a message and then later, that he is ready to share it. For instance, Mark had shared written guidelines with his professors on how to facilitate a discussion in a class where someone uses a communication device with suggestions including giving Mark some questions ahead of time so that he can prepare an answer, or allowing Mark to give a one-word answer that the professor can build on. Another solution Mark and his family came up with was to build a switch-activated LED light strip mounted right next to Mark's wheelchair's head rest. The LED light strip was made by D. T explained that Mark turns on his light to indicate he is ready to participate: "He turns it on to alert the professors he has an answer, [the light] is currently acting as his own sidekick when he is in class." The light was a nonverbal way to call for attention and indicate to others that Mark was ready but, it did not allow for more nuanced communication. We decided to explore how we could signal other social cues with motion to support Mark in managing turn-taking in group conversations.

5.3.2 Definition: Goals and Sidekick Properties

The design team met to discuss specific properties the sidekick should have to best support Mark's interactions and conversations with different partners. In this section we report on the identified goals and possible useful scenarios in which Mark envisioned using the sidekick. To scaffold our definition phase, the HCI researchers selected specific properties to discuss covering the function, social factors, and aesthetics surrounding the sidekick: (1) what conversational goals should the sidekick support? (2) How would the sidekick be controlled? (3) How should the sidekick be introduced in the conversation? (4) What should the sidekick do while it is inactive? and (5) What should the sidekick look like? We used these discussion points as probes to imagine different possibilities and identify design constraints.

Defining accessible sidekick controls. We learned that the preferred mode to control a potential sidekick would need to involve Mark's head-switch. The HCI researchers had brainstormed a series of controllers to discuss with Mark ranging from manual inputs to automatic sensing mechanisms, but when discussing these ideas, we learned that he had already tried and discarded many of these input modes before. T explained that Mark has mixed muscle tone due to his athetoid cerebral palsy, so other input modes that require motor precision and repetition of controlled movements such as eye tracking, facial gestures or foot pedals are not accessible.

"We have tried a lot of access points; can I go over them? we tried the elbow; we tried the knee. We tried some things with [Mark]'s hands... he can go in one direction but can't retreat from that direction so if he were to get his hand out here it might stay there and then but really, he needs to [bring it back] to release it as a switch. [Mark]'s most functional area for selecting is his head."

We also learned that gesture-based input could be tiring. Mark and his family had also tried a system that comprised of a wearable headband with electrodes that could sense winks and specific facial grimaces that could be detected and help Mark with accessing his communication device. They were trying to use it as an alternative to eye gaze but it was impossible to find a consistent facial movement. Having to do repetitive gestures was also physically taxing. Mark currently has two head-switches, one to control his AAC device and another one to turn his light on and off. Mark stated this was already the maximum number of switches he desired so he would prefer to be able to use the same switch he uses for his light, for the sidekick. So, we integrated the sidekick to be controllable by one of his head switches.

Sidekick interactions and inactive state. It was challenging to talk about sidekick properties and interaction without having a physical model of what the sidekick could look like. Mark expressed that the word "prototype" the HCI researchers kept using to describe a possible sidekick was not completely clear to him so HCI researchers clarified they meant a model of what

the sidekick could be. Clarifying that nothing was set in stone yet allowed us to freely explore the possibilities and talk about the constraints of each. We discussed how we imagined a possible sidekick would behave starting from what it would do when it is not in use and how partners would discover it. Prior to our meeting, Mark had answered that similar to his head light, the sidekick should always be present but not always active to prevent Mark from having to retrieve something that needed to be taken off and put on. We also had imagined that the sidekick could be hidden and it should appear suddenly when needed but this also begged the question as to how we could execute this mechanically. Mark indicated he would prefer for the sidekick to remain in its position when not in use. Mark shared that he would not know where to store a sidekick so he assumed it would work best for him if left on, while remaining ambient. We pinned this in our discussion and revisited it later asynchronously once we converged on a sidekick form factor.

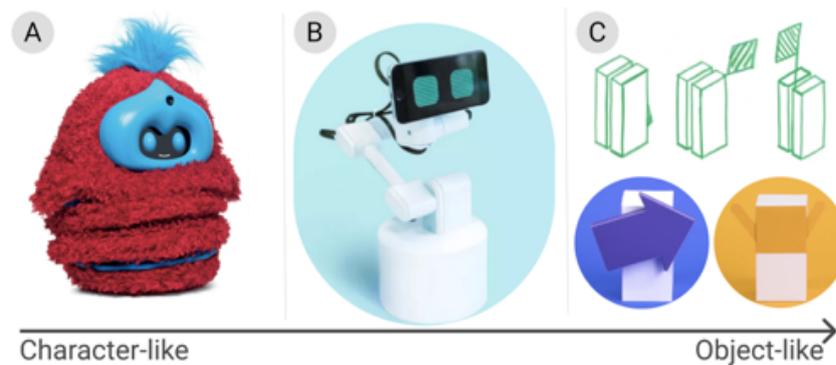


Figure 5.2: Expressive robotic objects ranging from character-like to object-like. (A) Tega, a robot emotive companion [164] image credits to Bruce Peterson, (B) Emoto, an emotive AI sidekick made of a robotic stand that acts as a body for your phone [19], and (C) Paper Signals, image credits to Google, animated objects that track and display data through embodied motion [18].

Physical appearance and placement. Mark mentioned that he wanted the sidekick to have a smile, but he was unsure about what he wanted it to look like. We conducted a literature review on expressive robotic objects that had smiles or faces and used the images gathered plus other expressive objects. Some of the images we used are shown in Figure 5.2. We shared these different expressive robotic objects examples using Padlet, a digital online collaboration board maker that Mark recommended for its accessibility. We placed each image in a way that Mark could comment below it asynchronously and use thumbs up/thumbs down to prioritize the favorite ideas over ones that did not resonate with what he had in mind. We also allowed a space in the collaborative board for sketching, and for labelling where the sidekick could potentially be placed on Mark’s wheelchair. While Mark did not sketch, he indicated preference for the flag-looking object from the Paper Signals project [18] that showed a flag raising up from a box-like container (Figure 5.2, right). Mark also indicated that it would be best to place the sidekick to one side, close to the AAC device, instead of other alternate options that included using the space in the back of Mark’s head-rest. Once we decided on the sidekick’s form the HCI researchers started prototyping, as described in the next section.

5.3.3 Development: Ideating and Prototyping

Once we decided on a flag-like form factor, we started exploring different possible motions. The HCI researchers created a low-fidelity prototype using a popsicle stick, paper, a servo motor, and an Arduino to showcase a variety of motions and share them with Mark and his family in a video. Figure 5.3 shows some example motions. The motions included: (1) rise and hold (to call for

attention like raising a hand); (2) home position (to demonstrate the idle state); (3) rise and wave (coming up and moving forward showing enthusiasm or agreement, like nodding yes, or calling for attention); (4) there-there motion (moving from 90 to 180 degrees slowly like saying “calm down,” or “It’s OK”); and (5) the metronome motion (moving from 0 to 180 degrees and then back while Mark is composing a message to show something is in process).

The HCI researchers also modeled a 3D flag-like object to ground discussions of what could be further developed in a high-fidelity prototype (Figure 5.4, A). We shared the video with the motions and the 3D model prototype with Mark and his family, and met to discuss the device’s development further. The HCI researchers focused on learning the sidekick’s size limitations and placement constraints, according to Mark’s needs. To do this, we carried out a conversational, spatial brainstorm in which we relied on shared visual information to ground our discussion. The HCI researchers shared the 3D model they had created and Mark and his family pointed to different places on his chair, while rotating their camera (used during video conferencing) around to facilitate our understanding of the space available and a potential mounting place.

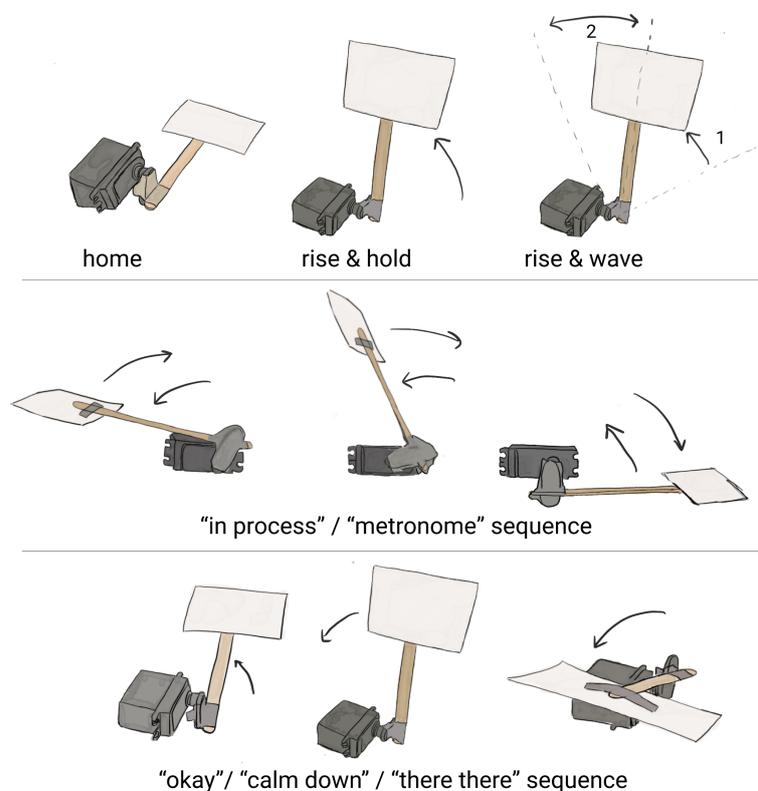


Figure 5.3: Low-fidelity prototype example motions

Constraining the design space. Understanding technical limitations and Mark’s access preferences was key in facilitating our discussion and making decisions about the sidekick. As we brainstormed different possibilities, D and T reminded the HCI researchers about Mark’s AAC device features and limitations. For example, Mark’s AAC device has Bluetooth and infrared remote-control capabilities that can be used to interface with a computer or another device. Nonetheless, the infrared and Bluetooth are not always reliable. T shared that the infrared control worked better than Bluetooth but requires a specific receiver that Mark only uses when he is working at a dedicated desk. It was also important to have a conversation on how Mark prefers to use his AAC device: *“Some people do use the computer side of [the AAC device]; Mark does not. He does not like shifting over to it because having the open computer where you have the language software on one side and the computer software on the other side, it slows down the language side. That is just one reason.”*

This conversation reaffirmed our decision to use the head switch as the way to more practically control the sidekick without needing to worry about wireless connectivity issues. These led us to a discussion about how we could enable triggering specific separate motions if we only had one head switch as our input. Our conversation turned into understanding the technical limitations of our envisioned sidekick. Could we built it to move only while the button is pressed? Should it stop once Mark released the button? Should it be voice activated too? Mark expressed he did not want to complicate things too much, by saying no to different ideas about using voice to activate different motions, and having the sidekick rotate to show different colors that could mean something. Such features would add additional learning and work to Mark's daily processes by requiring him to remember numerous sidekick states and controls.

Motion, a new material to explore. D has experience tinkering around Mark's wheelchair. For example, he built the LED light. However, D had not considered using motors before and was surprised to find how dramatic and descriptive even small motions could be. D explained how Mark can use his light to say yes and no, two light flashes for yes, one for no, but with the motor the object could move in a specific way that means something else, even when being triggered by the same switch. "When you showed the one you showed in the video, it was quite dramatic. It was almost like waving. So, he is like waving toward himself to get somebody's attention like saying "I am ready now"." When watching the example motion in the video we discussed how the different motions could be seen differently from different angles. For example, the metronome motion is more understandable from the side than from a front view. We decided to add a second degree of freedom that could support adjusting the sidekick to be visible at different viewing angles.

We thought that the desire to have a flag-like object could mean Mark might have been interested in adding a message to it. However, during our conversational brainstorm we realized Mark and his family considered that motions themselves were enough. D suggested extending the part attached to the motor a couple of inches higher, and removing the rectangular flag face area as it would probably be more practical, less vulnerable to wind, and make the sidekick smaller in size. Mark agreed and shared how he would like to change the square-like looking attachment to something like a popsicle stick and having the possibility of adding some extension to it in the future. D agreed: *"If the part that attached to the motor, that stick, is just extended a couple of inches higher, that is almost enough of a visual cue, if it moves up and down. I don't even know if it needs a rectangular area at all."*

Making use of the space available and contrasting colors. We also learned how having a very large sidekick could interfere with Mark's transfers out and in of the wheelchair. It was important to stay within Mark's chair perimeter to avoid obstacles. Mark had an existing mount for his wheelchair that could be used to place the sidekick right to one side of his AAC device without blocking his face or his line of sight. T and D offered to drop off the mount so that we could fabricate the sidekick around it. We decided that a sidekick with an approximate total height of 6 inches would be more than enough to be seen. Mark's family also suggested using a contrasting color to Mark's wheelchair and his accessories which are mostly all black. We decided then the sidekick would need to be a color that could stand out as D suggested: *"I saw the video of the thing moving. That is going to catch people's attention. You could have just the stick and people would see that. A bright-colored stick would do the job."*

5.4 Prototype Implementation

We implemented the ideas and feedback from our conversations into a fully functional sidekick prototype. We designed a mechanical structure made of 3D printed parts which is actuated by two micro servo motors with a range of motion up to 180 degrees of freedom each (Figure 5.4, C). The arm has holes that fit M3 screws and allow for easy mounting of other desired extensions. The device is connected to a metal piece that enables mounting on the wheelchair. The electrical design uses a METRO Mini 328 microcontroller from Adafruit industries and a custom PCB



Figure 5.4: A) Early prototype 3D model; B) Front view and C) side view of fabricated sidekick attached to wheelchair mount.

that connects the adaptive button and power lines to the servo motors and to the microcontroller. The device has a micro-USB connection which brings power through an external battery pack or directly from the AAC device’s USB port. A mono audio jack adaptor on the device serves as the connector for the adaptive button. The sidekick’s gestures were authored using the Arduino Servo library that allows setting motors to specific positions. To support others in creating their own sidekicks, we have open-sourced the 3D designs, software, and electronic schematics here: <https://github.com/Svsquared/AAC-sidekick>.

5.4.1 Initial programmed motions

Following video exchanges via email we programmed 3 motions for Mark to test including Mark’s preferred motion: (1) rise, pause and wave to call for attention; and two additional motions to explore and probe for ideas: (2) a return to home motion rotating the sidekick slowly inwards, and (3) pointing outwards as if pointing to an object nearby. The sidekick also included an intro motion sequence to signal being on and receiving power: once plugged in the sidekick would turn on and move to center itself. Each motion was programmed to be activated according to different button press durations: one fast button press triggered the rise up and wave motion, a press lasting about 3 seconds moved the sidekick back to home position, and with a longer press lasting about 6 seconds the sidekick rotated to point outwards. Each of these motions began with a “preamble” sequence which would bring the sidekick’s arm to the front and center of the device. This preamble was intended to capture people’s attention before the sidekick carried out the main motion.

5.5 Delivery: Testing and Iterating

We delivered our 3D printed design for a “test run” in which Mark would try the device for a couple of weeks and provide feedback. After trying out the prototype, Mark and his family came up with a new motion they wanted to use, called timer. Mark often says “can you hold on a minute please” to indicate he would like to say something and needs a minute to compose his message. The idea of the timer motion is to make the sidekick’s arm act as a timer that moves from side to side at a pace of 6 degrees per second, that is 30 seconds moving from left to right and 30 seconds right to left for a total of one minute side to side (Figure 5.5 and 5.6). Another suggestion was to add the word “typing” on the sidekick’s arm to further clarify the message.

Mark also found that the “preamble” sequence before each motion was more confusing than helpful in capturing people’s attention so we removed the preamble entirely. We also decided to drop the pointing motion as it was rarely used. After some iterations and feedback through virtual meetings we finalized a version of the sidekick that Mark would use for a longer period of time. The final version had only two main motions—the timer motion lasting 1.5 minutes total and the wave motion. The timer and the wave could be activated by a fast click and a 3 second

press respectively. Clicking the head button again during any of the motions stopped and reset the sidekick immediately (Figure 5.6), a key function we identified during the testing phases as it could help stop the sidekick in case of an accidental press of the head button or to stop the prolonged timer motion.

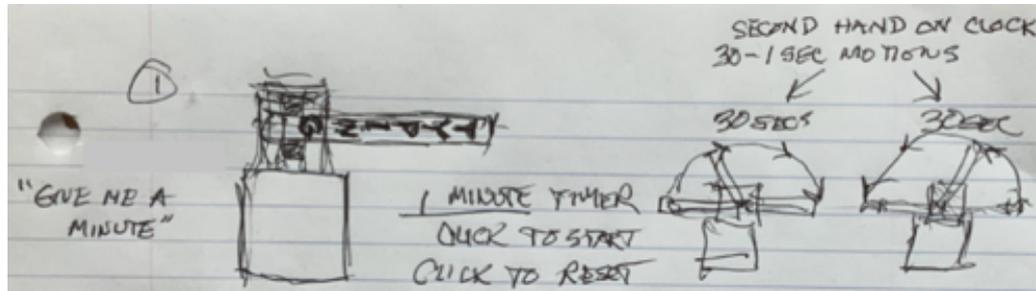


Figure 5.5: Hand sketch produced by D explaining the new timer motion.

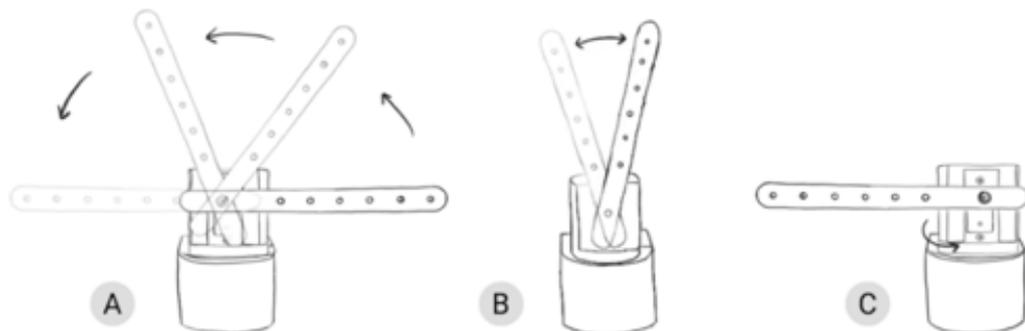


Figure 5.6: A) The timer motion moves 30 positions right to left and then 30 positions left to right, advancing 6 degrees per step; (B) The wave motion waves back and forth 6 times; and (C) the reset sequence centers the base and lowers the arm.

5.6 Evaluation: Use and Impact

Evaluating the sidekick for two months enabled us to understand if and how it supported Mark’s communication in different real-world contexts and with different communication partners. Using the sidekick for a long period of time also provided us with ideas for improvements and allowed us to reflect on how the day-to-day tasks (i.e., technology set ups) and different communication channels (virtual or in-person) impacted the sidekick’s effectiveness and use. We analyzed the diary entries provided by Mark, we graphed Mark’s circle of communication partners, and we met regularly to discuss how the sidekick was used and how it worked according to different interactions. We were interested in understanding if other people understood the sidekick’s purpose, if the sidekick enabled Mark to participate more in conversations, support him with turn-taking and show others his intent to contribute.

We collected 11 diary entries over three weeks. We asked Mark for information on any activities during which he used the sidekick and we asked him to rate its usability and social factors including: “The sidekick worked as expected”; “it was easy to use”; “It helped me achieve what I wanted”; “it was more distracting to me than it should have been”; “partners noticed the sidekick”; “partners understood what I was conveying when using the sidekick”; and “partners seemed distracted by the sidekick”. To better understand Mark’s relationship with the communication partners he interacted with during the evaluation period, we interviewed Mark and created a circle of

communication partners (Figure 5.7). His partners included family and friends. His three family members, T, D, and P, as well as his friends, acquaintances, service providers, and the new communities he is building through his advocacy work. The partners reported in the diary study are lightly bolded in Figure 5.7.

The sidekick was more useful during face-to-face conversation as compared to using it in virtual meetings. We also found close communication partners did not need the sidekick but understood its purpose right away. We also learned that there were some unexpected barriers to the sidekick use like remembering to make sure it was plugged in to power and to the head switch. We present our main evaluation findings next.

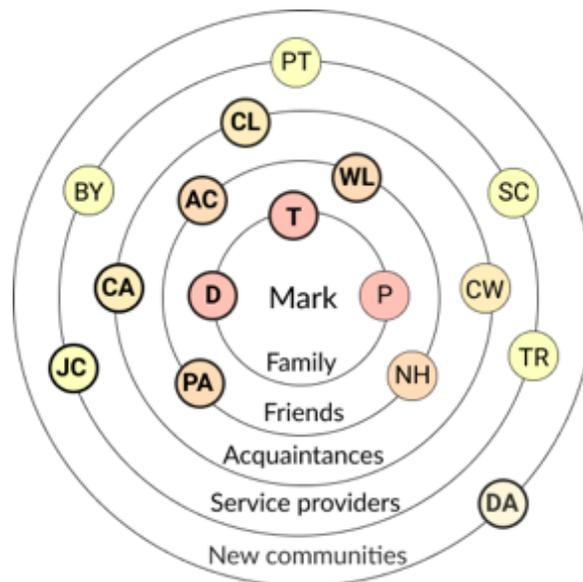


Figure 5.7: Parents (D, T) and brother (P) are included in the family circle. Friends include past aides (PA), AAC group friends (AC), music camp friends (WL) and neighbors (NH). Acquaintances include co-workers (CW), HCI collaborators (CL), and current aides (CA). Service providers include a job coach (JC), physical therapist (PT), Supports Coordinator (SC), and an assistant’s agency (BY). In the outer-most circle a disability rights community (DA) has recently started to get to know Mark.

5.6.1 Sidekick Use and Performance

The sidekick was reportedly used 7 of 11 days, summarized in Table 5.1. The sidekick was used between 2 to 5 times per conversation on 6 days and it was used between 6 to 10 times on one day when Mark was talking to his friends from an AAC conversation group. We learned that it was easy to forget to plug the sidekick in and this was the main reason for days of nonuse. The sidekick was already mounted on the wheelchair everyday but it needed to have the head switch cable connected to it to be operable, as well as making sure it was connected to the AAC device for power. One improvement suggested by T was to make the sidekick integrated with the device, in a way that it could be controlled with the AAC device and it could not need external cables.

Mark reported that 7 out of 7 times others noticed the sidekick during his meetings. The sidekick was easy to use and worked as expected most of the time, although for one meeting with his job coach, the sidekick worked sporadically – requiring Mark to press the head switch twice or for a little longer to make sure it started the motion. Even though the HCI researchers could not replicate the latency problem we reflected on the importance of making sure the sidekick was sensitive enough to Mark’s press frequency and style, as it was evident from his report that the

microcontrollers internal delays were not allowing the button to be sensitive enough to his desired rhythm of use on some occasions.

Mark reported that the sidekick helped him manage turn-taking when talking to his AAC group friends and when talking to a group of his current and past aides with whom he is close. The sidekick did not really help with turn-taking that much for his other conversations. The sidekick did not really increase Mark’s participation in meetings but it was helpful in reminding partners to wait for Mark’s response and was used frequently to replace his preprogrammed message: “can you hold on a minute please?” In general, having access to a motion-based AAC sidekick enabled a short-cut in communication, allowing others to understand typing is happening and allowing Mark to not have to verbally say “one moment please” with his AAC device.

Table 5.1: List of meetings reported with diary entries over a period of three weeks.

ID	# of partners	Relationship	Medium
A	2	Past assistants (PA)	In-person
B	5	New and past assistants (CA, PA)	Online
C	5	New and past assistants (CA, PA)	Online
D	15	Music group friends (WL)	Online
E	5	AAC group friends (AC)	Online
F	1	Job coach (JC)	Online
G	15+	Disability rights advocates (DA)	Online

5.6.2 Close and New Communication Partners

The familiarity with AAC and also with Mark’s communication style is what makes close communication partners skilled enough to not need the help of the sidekick, nonetheless they understood its purpose right away. During a drive-in visit to some of Mark’s past aides who have known him for more than 4 years and have become friends (meeting A, Table 5.1), Mark shared they asked what the sidekick was and he demonstrated how it worked and they “got it right away.” “If they know you well, they do not need the extra help. It is sort of like a novelty but not absolutely necessary because those folks are waiting; they are patient and they are waiting. If they see Mark’s body language that he is writing they know Mark is typing.” Nonetheless the sidekick seemed useful when Mark talked to the same group of past aides’ friends and a group of current aides (meeting B and C, Table 5.1) due perhaps to the fact that this meeting was online and it had more people in it, which could make turn-taking a little more challenging.

The sidekick was also useful when Mark talked with AAC friends who are familiar with the workflow involved in being an augmented speaker. Mark shared that some of them had told him they wanted one for themselves. The sidekick also seemed useful in one occasion with Mark’s mother at home. Mark was in another room and he triggered the sidekick to call his mom, she heard the sound and caught it moving and realized Mark was calling for her attention.

We hypothesized that the sidekick would be more useful with unfamiliar partners, allowing Mark to better regulate turn taking dynamics with people who had not met him before but we learned that this was hard to measure in an online setting as unfamiliar partners needed to first be introduced to AAC and Mark’s communication style online, which brings new constraints, in addition to having to interpret the sidekick.

5.6.3 Better in-person

The sidekick was originally designed to support face-to-face conversation but due to the COVID-19 pandemic, the group conversations that Mark had during our evaluation period were mostly online with a couple of exceptions for when he did some drive-in visits to friend's houses or used it at home. Mark shared that the sidekick helps more with turn-taking in person than it did online. The visuals a person can get during a video call can be limited by the number of people on the call, the positioning of one's camera and people's attention to the screen. This was the case when Mark was talking to the disability rights advocates. Mark has found that the best alternative when meeting with large groups on video is to use the chat instead of the sidekick; T elaborated on his answer:

“Like today he was on a call and there were like 30 people on the call. So, he is using sidekick, your picture is small and sometimes there is not a single facilitator but the speaker is shifting around. and it's not like there is not anyone in place who is going to call on you. And in that situation, you may be better off using the chat. I know he had to do it today. He said, 'I have a question.' And said it in the chat. The visual for the zoom call in those situations where there is a lot of people and they are not 100% used to AAC, they are not going to have a reaction to the sidekick or even know what it is.”

Mark also explained that people not seeing the sidekick was also due to his camera's position. The family has worked on positioning the camera to make sure Mark's face is shown without cutting off the sidekick but it is sometimes challenging to get a good set up for every meeting. *“I think it is fantastic in real-time with real people. You know, face to face. There is no question. They are seeing it. It's here, you know, but in the digital meetings, it is more difficult.”*

Surprisingly we did find that using the sidekick online was useful with acquaintance or “mid-circle” partners who were familiar with Mark but who were not as skilled as those partners in the first two closest circles. For example, when talking to the design collaborators of this project on Zoom, Mark effectively used the sidekick to help others pace themselves and their questions. The sidekick also supported Mark to talk with his group of friends and current aides, and when talking to his augmented communicator friends during his AAC group by helping Mark show others he was composing a message.

5.7 Discussion

As a case study, this project demonstrates how motion can support AAC interactions and conversations. We argue that motion-based AAC achieved through expressive objects is a promising new communication modality to continue exploring. Our work revealed specific ways in which a physical expressive sidekick supported an augmented communicator's interactions, uncovering additional opportunities for future work. We also learned lessons on how to scaffold co-design activities to collaborate as a team and envision and develop a new technology.

5.7.1 Benefits of aided nonverbal communication

People familiar with AAC understood the sidekick's purpose right away. Experienced communication partners did not need the sidekick to communicate with Mark as they know how to read Mark's body language and they know to patiently wait. Close partners know how to make use of the rich multimodal communication already happening, understanding when Mark is composing a message or he is assenting or dissenting with his blinks. Nonetheless having a physical and visually salient device can make Mark's communication intent much clearer, especially for conversation partners who do not know where to place their attentional focus. By augmenting embodied

interaction through an external physical device, we are making it evident to other partners that an AC is conveying an action that has communicative intent. The sidekick’s timer motion leverages a familiar cue that can support unfamiliar partners to understand that Mark is typing and he needs time to compose his message. Similar to other modalities such as screen-based emojis or LED lights [146], motion can grab immediate attention but it can also convey precise messages in an ambient, peripheral, and spatial fashion to overcome display resolution limitations.

Our evaluation approach allowed us to test the sidekick in-the-wild but with some data limitations as we could not control the types of conversations that occurred. While we have not yet conducted a rigorous comparison of sidekick use with familiar and unfamiliar partners, we learned that the sidekick had the potential to make it easier for Mark to communicate specific things to unfamiliar partners, such as “I am typing,” “hold on a minute” (while the sidekick moved), and “I am ready to participate” (once the sidekick stopped moving). But we learned that unfamiliar partners did not incorporate the sidekick into conversations seamlessly; it was actually mid-level conversation partners who incorporated the sidekick the most. Due to the nature of the conversations being mostly online we learned that the sidekick might be better introduced to new partners in this context with a message to tell new partners what they should focus on. For example, Mark could say “the red item on my device indicates that I am typing.” Thinking about how different partners in Mark’s circle have different communication abilities poses the opportunity to explore future communication technologies that are specific to partner types. For example, systems using more complex nonverbal gestures can be used with close partners who could spend more time learning a new system or a new language, or even providing content input [46].

5.7.2 Barriers to long-term use and AT development

The use of assistive technology (AT) has been reported to be low even when people have access to AT [38, 116, 122]. The reason for this is usually connected to usability barriers and social acceptability—social barriers that impact AT use [139, 140]. We identified some barriers to using the sidekick device long-term related to having to remember to plug it in and making sure the camera was set up at a right angle, both related to the daily set up routine. For instance, Mark told T: “We are going to have to remember to plug it in,” illustrating that having to add extra steps into the daily technology set-up is not trivial and can be a barrier to using a new system. AT should aim to be mostly integrated to existing technology use, but this can be challenging when current AT systems such as AAC devices are not open to developers to build on and integrate new features in. Other developers have encountered the same limitation, the lack of a complete AAC functionality stack into which new developments can be built [46]. During our development phase we spent a considerable amount of time learning more about Mark’s AAC device’s capabilities and about his workflow – how he preferred to use a separate computer for Zoom and keep his AAC device mainly for communication. There were a lot of “unknowns” regarding how compatible his AAC device was to other peripheral devices. We decided to go with the stand-alone, head switch operated sidekick to make progress and make something work but this tension illustrated the boundaries of individual co-design, making it harder to recommend how bespoke technologies like this one can reach a wider audience.

Another factor that can impact AT use is access to proper maintenance of a device or troubleshooting over time. Maintenance of bespoke technology designs must be an important factor to consider to ensure it is used. T brought up this important point: “If this turns out to be a really helpful thing, then where do we get the technical support to keep the prototype functioning?” To address this, we open-sourced our design for others and also connected Mark to local makerspaces and volunteers working on making bespoke open-source assistive technology. The HCI researchers will continue to provide support for this device but having a long-term plan in place can make sure that maintenance is possible beyond the HCI researchers is also crucial. Though this tension of impact and maintenance is unresolvable with this project, we found that empow-

ering Mark by making the design open and by keeping clear documentation is a starting point to ensure that end-users know what to ask for when seeking technical support. Future work may look at how existing online communities developing Do-it-Yourself AT [69] can continue maintaining and expanding motion-based AAC solutions.

5.7.3 Accessibility of the long-term co-design process

Carrying out our co-design collaboration over an extended period of time was beneficial, allowing us to develop valuable relationships and reflect on the process through accessible iteration and prototyping as found in prior work [40, 52]. To collaborate with Mark and his close communication partners T and D, enabling multiple feedback channels via email, video, drawing, collaboration boards and video chats was key in helping us carry out the co-design process. For example, we were able to take up the specific tools such as the Padlet ideation board that Mark recommended. Often, co-design is engaged with the assumption that designers provide resources during in-person workshops. We realized after the fact that we engaged privileged skillsets to move co-design online. While we may have been able to provide institutional access to digital tools and we provided the physical components necessary to prototype sidekicks, we took for granted the technical skill required to join us on these platforms. While Mark and his family were tech savvy, we did find that our explicit conversations with Mark about which communication tools would work for him were still important for us to have effective design sessions; co-design concerned the process, not only the prototype. As we recognized open communication about tools and techniques seemed to be a positive starting point, there is a need to explore co-design of bespoke technology during remote collaboration that can leverage different tools. For example, we learned a lot by dropping off preliminary prototypes with Mark and having the spatial conversational brainstorming sessions, where over a video call we learned the feasibility of different possibilities in fitting onto his wheelchair and into his overall space.

We also learned that less was indeed more. In co-design and ideation more generally, designers diverge to generate many different possibilities. Early in our process we came up with many drawings and ideas of where the sidekick could be placed and how it could be accessed and controlled (voice control, facial gesture recognizer, multiple buttons, etc.). We wanted to think of anything as possible as designers but as T and Mark shared, they have tried a lot. Harrington et al. critique “blue sky” ideation with critical race theory, that the underserved, black communities they worked with know what types of structural changes might enable access and ideating things that won’t come to fruition can widen gaps between what different co-designers perceive of as ideal. Instead, Harrington et al. and Bennett et al. recommend understanding stories and rich accounts to recognize the knowledge and labor co-designers have already expended [10, 61]. In Mark’s case, he and his family have already done the early exploration of figuring out ways to make communication easier and finding the right access modes. They have actually spent a lot of time and worked very hard, getting creative about many possible points of Mark’s body to use for access. They wanted the HCI researchers to know what isn’t going to work right away, and that it is very important to listen. We recommend that co-design actively incorporates ‘what doesn’t work’ into design sessions. Further, in-depth listening to understand co-designers’ process of iteration and ultimately determining the options that were not feasible was useful in helping us to not replicate these mistakes; in other words, lists of what not to do are helpful, but engaging the iterative everyday design that got to that point gave texture to the bad ideas that kept us from developing similar past solutions and unusable possibilities such that we did not attempt to replicate them.

5.7.4 Limitations and Future work

One limitation of our design process was that the hardware design iteration was done mostly by the HCI researchers as changes required 3D printing and 3D modeling and the motions were pro-

grammed directly to the sidekick's microcontroller. Next steps should include making the sidekick more robust to allow customization on the go. A future sidekick platform that enables authoring gestures via remote control could facilitate motion customizations by the user to produce changes on the go and explore new combination of gestures and without the need of HCI researchers in the loop. This flexible customization platform could also be furthered developed to provide ways to visualize other sidekick forms. Second, our co-designers had access to a lot of resources and had worked together for a long time to augment Mark's communication. As such, they quickly integrated into the design team. Future research should concern activating co-design that may support co-designers with different resources and experiences with DIY. Finally, making the sidekick more integrated with the AAC device in a way that it could be controlled by it is a clear opportunity for improvement and future work.

5.8 Summary

We explored how motion could support augmentative and alternative communication by co-designing and evaluating a physically expressive sidekick object with and for Mark. Using bespoke sidekicks that move in physical space as a form of aided nonverbal AAC can provide augmented communicators with an additional expressive output that can support them in managing conversation dynamics. By working closely with Mark and his family we also learned about the possible barriers to integrating a new device in daily life and in sharing our lessons, we look forward to future work in improving tools that support developers in building for AAC.

Part III

Agency through Language

Chapter 6

Discovering Opportunities for Language Technologies

When conversation shifts topics quickly and a person speaks about a previous topic that happened minutes ago it takes extra effort to understand what the comment is connected to as it may be initially understood as “out of context”. In addition, when AAC users respond to a question with a short utterance, it is hard to know what they are responding to if they have been asked many questions at once or if the group is talking about multiple things at the same time. In Chapter 3) we found that when conversations move while the AAC user is still composing a message, they produce missed opportunities for AAC users to participate. AAC users often erase their own comments after other conversation participants have changed the conversation topic [157]. One way to address the relevance constraint in conversation with NLP could be to provide ways in which an AC could address any past topic introduced in the conversation with enough contextual information as to make it clear to other partner’s on what they are talking about.

In this chapter we answer the research question, how might we use language technologies such as Large Language Models and natural language processing tools to support augmented communicator’s agency and support mutual understanding?

To answer this question I carried out two separate formative studies. First, along with my collaborators, I closely examined two conversations among AAC users, a close communication partner and an unfamiliar partner to understand when misunderstandings arise. Second, we carried out a two-day virtual participatory design workshop to explore the design space where language-driven technologies could help support AC conversations.

6.1 Conversation Interaction Analysis

In order to find out if AC’s might need additional context to their responses, a collaborator researcher and myself analyzed two different conversations among an unfamiliar partners and an AAC user and their caregiver. Each interview was around 30 minutes long. We tracked every response from the user as well as the conversation that led to their response in the conversation samples. We also looked at every instance when the AAC user nodded yes or no. We had previously labeled this data for our study on conversational agency (see chapter 3). In total we had 90 data points which consisted of the identified context the AAC user was responding to and the AAC user’s response. We mainly focused on the cases where the AAC user was conversing with the unfamiliar partner, or when all three people were speaking with each other. Thus, we did not consider direct questions from the caregiver towards the AAC user as data (i.e., Caregiver: “Would you like me to adjust the screen”; AC: Nods). However, every other response from the AC whether spoken or gestured was considered a data point. We were specifically interested in learning how we could extract the relevant context an AAC user may want to address in conversation, where the

relevant context could be located in time, and in what form was this relevant context presented in (whole phrases, nouns, verbs, other).

Each data point contained all the spoken words and gestures from the participants starting with the question and ending with AC's response. Within each data point we wanted to find the most contextually relevant information from the participants in order to ground the AAC users' responses. Finding this information is easier in some cases than in others. Consider a short direct question from the interviewer (Exp) to an AC:

EXP: How old are you

AC: Twenty-three

If the AC responds to this question without the caregiver talking, and the AAC user responds with their age, it is clear that the contextual information is simply the question asked. Now, consider the following more challenging example:

EXP: Do you use your device for social media?

AC: No.

CCP: [*Addressing Exp.*] Her?. Yes and no. This.. Her previous device, she uses as the main interface for her desktop computer at home that she does all her Facebook, email, Instagram... [*Continues explaining for over a minute and changes topic to talk about bluetooth connection*] ... So we kind of keep that separate from going out on media.

EXP: Yeah, that makes sense. It's nice that this connection allows the AAC device to be an interface to the computer and just like control it from there. Right?

CCP: right.

EXP: [*Changes topic*] And so you say you have used a device from before, was there any? was it similar to this one like was it cursor based? Or was it touch?

CCP: Direct access. We use a stylus on a lanyard that way.. [*Continues explaining for a long time (around 2 minutes)*] ... She uses both hands on it because her whole body is involved in her disability ... Using the Bluetooth joystick makes her look more natural and people can look at her while she is typing plus the fact that it is faster and more accurate.

EXP: Yeah. It looks more comfortable.

CCP: Oh very.

AC: I can. but I haven't explored the internet.

The AAC user's last response is responding to the original question at the beginning of the transcript. However, depending on what they said in their second response, they may have been responding to the more recent topic. In order to find the context, the entire transcript needs to be read. Therefore finding the most contextual piece of information in this example is more complex than in the previous example.

6.1.1 Analysis

Analyzing our interview data in detail allowed us to find patterns in the data to understand how we could extract a relevant portion of the conversation an AC was speaking to. We had two initial hypotheses: (a) that we could find what the AC was speaking about within past turns in the conversation and (b) that we could use the time at which an AC started typing as an indicator of where the most likely relevant context could be found. We asked the following questions: For how many AC speaking turns is there context available in the conversation? and how far away from the typing event is the relevant context?

All data points were analyzed by myself and a research collaborator and discussed to make sure we reached agreement on what we believed was the relevant portion an AC was speaking to. We used the complete conversation to determine this relevant portion.

6.1.2 Results

We found that for 100% of AC speaking turns there was context available in the conversation preceding the AC's speaking turn. We also found that 79.4 % of the times the context the AC is responding to was in the prior turn to when the AC started typing. The time at which an AC starts typing can be a strong signal that can help find the potential relevant portion.

Types of observed typing behavior

We observed three types of general typing behavior that impacted the location of the relevant context portion:

Hear and then reply: Often the AAC user waited for another speaker to finish their turn before proceeding to type and speak their contribution to the conversation (79.4% of the instances in our data).

On the go typing: In other cases (in 14% of the time) the AAC user types while other people speak and change their answers based on how the conversation advances. AAC users either speak to the same topic that was introduced when they started typing or erase their message and talk to a new topic that surfaced while they were typing.

In chunks: Other times (around 6% of the time), AAC users spoke to the current relevant topic through different speaking turns. That means that the context they are pointing to is their prior utterance and not in their partner's speaking turns. The relevant context is connected to their initial typing time and is elaborated upon more turns that follow it which are sometimes interrupted by other speakers.

While our conversation interaction analysis is limited in the sense that only two cases were thoroughly examined, we brought these examples and observations to a workshop with AAC experts to gather a wider understanding of the challenges around establishing mutual understanding.

Opportunities to support mutual understanding

In summary we found that the start of typing time could be used as an event marker to support an NLP application in knowing where the context in a conversation could be and how we could retrieve it to make it useful to an AC. Nonetheless there are still unknowns such as how we can leverage rituals between conversants (asking each other questions, helping each other achieve mutual understanding) to support AC's agency with NLP. Our preliminary data analysis was also done on limited data that only reflected an interview setting type of conversation. Next steps in this exploration with NLP needs more data-based examples and input from ACs on how this technology material could and should support their conversational agency in AC. In the next sections I describe how complemented this initial data-driven analysis with additional input from subject matter experts and ACs.

6.2 Interdisciplinary Perspectives on Language Technologies for AAC: A Participatory Design Workshop

We carried out a two-day virtual participatory design workshop to explore the design space where language-driven technologies could help support AC conversations. Participatory design workshops that are carefully designed to scaffold discussions among diverse stakeholders have shown to be productive in developing a congruent view of the problem space in areas of accessibility that require an interdisciplinary approach [20]. Participatory design seeks to bring multiple experts, including end users to design together and elicit new research directions for AAC through selected co-design activities [128, 129]. The participation of distinct stakeholders in the design process is valuable in enabling interdisciplinary design thinking for problem-solving [37, 121].

A participatory approach can help us avoid making assumptions about what technology solutions would bring value to users which leads to many products that are not usable and quickly abandoned [120, 86, 67].

6.2.1 Participants

Workshop participants (Table 6.1) included two augmented communicators (ACs), two computational natural language researchers, three communication scientists, and three HCI researchers. AC participants (AC1 and AC2) included two English-speaking men who used AAC due to cerebral palsy (AC1) and a spinal cord injury (AC2). AC1 used a Prentke Romich AAC device with the Unity Language software. AC1 used his fingers to directly press buttons on the screen of his device. AC2 used a Tobii Dynavox gaze-based speech-generating AAC device with the Communicator 5 software. Both AC participants used phrase-level voice output, speaking a phrase or word once it was completely typed. Both ACs participated virtually and had their camera on during the study.

Computational natural language researchers (N=3, 1 woman) included two professors and one Ph.D. student working in artificial intelligence (AI), natural language processing, language understanding, and generation for AAC, storytelling, and human-AI interaction.

Communication science participants (N=3, all women) included a certified speech-language pathologist (SLP) specializing in AAC for literacy and learning, a sociology Ph.D. student studying the multi-modal communication of children with speech disabilities who use eye tracking AAC systems, and a Ph.D. student (and former SLP) studying technology-mediated communication and miscommunication in robot-assisted surgery and AAC.

Human-computer interaction (HCI) researchers (N=3, all women) had expertise in accessibility, human-robot interaction, and natural language processing. The HCI researchers included two assistant professors and one Ph.D. student who led this work.

6.2.2 Procedure

The workshop was split into two sessions. On day one of the workshop, participants shared interaction-based challenges and brainstormed potential solutions through three writing exercises. On day two, workshop participants discussed all the challenges and solutions shared on Day One and further refined specific solutions.

Day 1: Sharing interaction-based challenges and solutions

- **Introductions:** All participants first introduced themselves. The workshop organizers introduced the goal of the workshop: to uncover opportunities for technology to support augmented communicators in participating fully in conversation and presented examples of existing language technologies.
- **Challenges in AAC:** AAC participants shared prepared statements of their communication challenges and experiences with all attendees. The workshop organizers also presented existing challenges in interactions between AAC users and non-AAC users reported in the literature.
- **Brainstorming Activity:** Participants were split up into two teams (Team 1 and Team 2) to carry out a structured brainstorming activity in a smaller group. Team distributions are specified in Table 6.1. Using a virtual board where everyone could upload and comment on each other's ideas, participants were instructed to individually brainstorm by writing their thoughts in connection to three specific prompts. We prompted participants to reflect and write independently about the following:

Table 6.1: Participants included expert augmented communicators (AC1, AC2) with more than 10 years of experience using speech-generating AAC devices; Communication Scientists some who were also certified speech-language pathologists (SLPs) (COM1, COM2, COM3), and technologists specializing in natural language processing and generation (NLP1, NLP2, NLP3) and Human-Computer Interaction and accessibility (HCI1, HCI2, HCI3).

Participant ID	Expertise	Background/Details	Team #
AC1	AAC user	Direct touch input, cerebral palsy	1
AC2	AAC user	Eye tracking user, spinal cord injury	2
Com1	AAC, Literacy and learning	Certified SLP, Postdoctoral fellow	1
Com2	AAC, miscommunications	Certified SLP, PhD student	2
Com3	AAC, Multimodal communication	Social anthropology and communication, PhD student	2
NLP1	AAC, Dialog Act classification	Computer Science PhD student	1
NLP2	Accessibility and collaborative language generation	Professor	1
NLP3	Story generation, human-centered AI	Professor	2
HCI1	AAC, Accessibility	Human-computer interaction PhD student	1
HCI2	Accessibility, Machine Learning, NLP	Professor	2
HCI3	Human-robot interaction, human-AI learning	Professor	2

1. A challenge they are interested in solving
2. A technology that would help an AAC user’s familiar or unfamiliar partners to be better communication partners. What would you like the partner to get better at?
3. A technology that would help the AAC user by either proactively knowing what the AAC user needs or by providing assistance whenever the AAC user actively requests it

Participants were given 7 to 10 minutes for individual brainstorming per prompt, 3 minutes to read and either comment, upvote or downvote each other’s ideas on the board, and 5 minutes to discuss as a group.

Day 2: Refining potential solutions

- **Revisiting ideas from Day 1:** All participants reconvened as a large group for an overview prepared by the workshop organizers about the challenges and solutions shared by both teams on day one.
- **Ranking challenges and solutions:** We asked AC1 and AC2 to identify their top three challenge areas they would prioritize. After they stated their selection, workshop participants discussed how different solutions connected to these challenge areas could be implemented.
- **Co-designing:** The workshop participants split into their original teams to discuss specific solutions that could address a prioritized challenge. Teams discussed the specific challenge they selected, their proposed solution to that challenge, and potential downsides to the proposed solution. They recorded these ideas in a shared online document.

6.2.3 Data collection and analysis

Video and audio recordings were collected during the two-days of the workshop. After Day 1, workshop organizers collected all the posts from the online board generated during the written brainstorming activity and organized them by common themes. These themes were phrases as “tools for X” where X represented the different challenges participants wrote about addressing. During day 2 this list of tools as well as a summary of the specific solutions (see Figure 6.1) people came up with during day 1 was presented to the whole group to continue the discussion. Each

Challenges	Opportunity areas/Participant solutions
Grammar & Pronunciation	Tools for message presentation <ul style="list-style-type: none"> • Auto. adjust grammar • De-link intelligence from grammar • Add pronunciation to words
Misunderstandings and Repair AC2 AC1 	Tools to speak to a specific topic <ul style="list-style-type: none"> • Topic Stack/Marker • Topic timeline <i>Team 2 idea</i> Tools to augment common ground <ul style="list-style-type: none"> • Marking repair by enabling word edit • Mind maps • Displaying communicative intent (asking, protesting etc.)
Accessing New Vocabulary AC1 AC2 	Tools to retrieve expressive vocabulary <ul style="list-style-type: none"> • Movie/pop culture phrases • Alternative word suggestions <i>Team 1 idea</i> • Vocabulary search bar • World events vocabulary • Routine-based vocabulary
Partner feedback AC1 	Tools to give cues and feedback to communication partners <ul style="list-style-type: none"> • Thought bubbles • Show typing status • Composition alert/flag • Interactive reminders • Computer moderator
Managing turns and timely participation	Tools to visualize speaking turns <ul style="list-style-type: none"> • Plotting typing events • Conversation threads Tools for adaptive AAC <ul style="list-style-type: none"> • Conversation-aware systems • Conversation follow-up

Figure 6.1: Challenge and opportunity areas uncovered in our participatory workshop along with names for solutions participants came up with. We highlight the challenge areas AC participants identified as bringing the most value to them. AC1 ranked *accessing new vocabulary* as his top choice and *misunderstandings and repair* as his second. AC2 was highly interested in tools to help him address past topics in conversation and ranked *misunderstanding and repair* as a priority area followed by *partner feedback* and *accessing new vocabulary*.

challenge area and specific solution was labeled with a letter and number to facilitate referencing it throughout the discussion. We used these numbers to keep track of the data.

On day 2, AC were asked to select 1 to 4 challenges or specific solutions they wanted to focus on for the rest of the workshop. We collected these rankings by placing them on a shared slide that displayed an axis labelled “More value to user” to “Less value to user”. We used this visualization to scaffold discussions about technical feasibility and support teams in ideating a solution to address the prioritized challenges. To capture information from the co-design team activity, we provided teams with a slide where they would write down specifics of their idea: (1) a description of the solution, (2) a description of what it solves, (3) the expected outcome (4) and a reflection of how the technology could fail. These categories were inspired in Storyboarding [155, 60].

After we completed the two workshop days, we carried out inductive qualitative research, moving from specific observations of topics discussed during the workshop to broader themes identified through thematic analysis [22]. These resulted on the broad challenges areas shown on Figure 6.1.

6.3 Challenges and Opportunities for Aided Support

In this section, we synthesize challenges (C1-C5) faced by ACs and other workshop participants as identified on day one of the workshop. We outline five different challenge areas (Figure 6.1) brought up by the participants and highlight each discipline's perspective regarding opportunities (O1-O5) to address such challenges.

6.3.1 C1: Grammar and Pronunciation

Augmented communicators (ACs) shared challenges related to grammar use and pronunciation which affected the presentation and reception of their spoken message.

ACs shared having challenges with using grammar when they are replying to a message in real-time due to time pressure and limitations with the language system in their device. ACs might omit using prepositions to save time and get their point across as soon as possible. In the case of AC1 who uses sequences of icons that turn into words (unity language system) using articles and prepositions requires having to navigate to his keyboard to type letters individually when the system does not correctly predict them. Both the time pressures and the icon-based AAC systems make using grammar harder for users which has social consequences like people perceiving AAC users as less competent. AC1 shared that people expect AAC users' phrasing of sentences to be grammatically correct and that when that expectation is not met, people might struggle to understand the message or assume he is not able to communicate effectively. When introducing himself to the workshop participants, AC1 wanted to clarify his use of grammar and establish his competence:

"the way my speech device is programmed, my sentences are not always grammatically correct. This is not because I'm not able, but rather is the limitation of the system I use." - AC1

Following AC1 testimony, COM1, COM2, and COM3 discussed how there is a societal pressure to use correct grammar that extends to AAC users even when the AAC device may make this difficult and the expectation of perfect grammar should not be the same as with spoken or non-AAC written text.

Another challenge that surfaced was achieving correct pronunciation. Digitized voice software used in AAC devices often mispronounces proper nouns like last names and lacks a way to control the speed at which speech is read and intonation. In order to make sure text-to-speech engines get the right pronunciation of proper nouns like last names or read-out messages with the correct cadence like a long phone number, AAC users have to manually insert extra letters or punctuation to their messages. AC2 shared that he re-spelled his last name so that his AAC device could read it out loud correctly and he could correctly introduce himself. Having to rearrange letters or add extra punctuation requires significant planning and effort. AC2 explained via a written message on the collaborative online board:

"I use phonics when I am typing. I go over all of the words I have set up in my word prediction so that the wording sounds correct."

AC2 also shared that if pronunciations are not fixed they can easily lead to misunderstandings or embarrassing moments that distract others from what he is trying to say.

O1: Opportunities for Message Presentation Support

Participants identified a social opportunity to help communication partners decouple grammar from their perceptions of ACs' intelligence, and a technical opportunity to help ACs produce statements with their intended grammar and pronunciation:

De-linking intelligence from grammar use in AAC. HCI1 and HCI3 mentioned that having a tool that could automatically suggest or adjust grammar to AC1's utterances could help. Communication science researchers (COM1 and COM3) agreed but insisted that using correct grammar was not only a technical challenge but a social one and that the focus should be on finding a way to make people less demanding of other people by making them more aware of AAC technology limitations. Using proper grammar with an AAC device is not expedient, but people (unfamiliar with AAC) think less of speech that is not grammatical.

"...like parents asking perfect spelling from children. They already understand each other but parents expect perfect and make them erase and correct. Often times there is an expectation that messages be "grammatically correct" in the way that we expect written text. Especially for children and adolescents. Would love a solution to support partners in accepting messages." - COM3

Making pronunciation easier. HCI3 suggested having AAC devices source correct pronunciations from a voice bank containing crowd-sourced recordings of unusual words, making the connection to existing name pronunciation recordings available on some personal websites and university student lists. Participants discussed creating a notation system ACs could use to indicate the right pronunciation of a word on the AAC device.

6.3.2 C2: Misunderstandings and Repair

Misunderstandings are common in every type of conversation [130, 133] but when they occur among AAC users and non-AAC users, repairing or clarifying these misunderstandings can be very challenging. Communication scientists explained that when misunderstandings arise, individuals commonly use the sequence of spoken conversation to determine where the misunderstanding occurred [133] but because AC's utterances are often a few seconds or minutes delayed the relevance of their message is often lost.

Additionally, ACs tend to reply to questions in conversation with short utterances that may appear ambiguous or irrelevant when the conversation sequence is lost due to message composition delays. Speech is less ambiguous when it is *in context* of the conversation (e.g., A: "do you like my dog?", B:"no, he bit me"). Imagine communicator B does not say "no, he bit me" until after A says something else like "do you want to go to lunch with Sam?" it could be confusing. People often follow the sequence of spoken conversation to determine the relevance of a speaker's message. Due to AC's utterances being often a few seconds or minutes delayed, the relevance of their message is often lost leading to confusion. Usually, individuals avoid misunderstandings by adding additional references to their utterances (e.g., "remember when you asked me about your dog? well...") [30] but adding additional contextual references that remind your partner about a past topic requires additional effort to an AAC user (i.e., more words to type) which means the message they intend will be even further delayed. As such, it is common for AC's responses to lack contextual references when having to respond fast in real-time making them ambiguous if the context is hard to interpret or remember.

Ambiguity about what the AAC users' utterance intended communicative action is (e.g., asking a question, expressing doubt) or their communicative stance towards the topic (disagreeing, agreeing, protesting) can also lead to misunderstandings. In addition to identifying "what" is being discussed by the user, or the topic they want to address, COM3 suggested that knowing the user's communicative stance could be useful, that is the attitude they have towards a topic. There are several different stances to be taken on a topic, you can ask a question about it, comment, react to it, or reject it. Communication scientists (COM2) explained that oral speakers tend to vary their intonation to stress words or use disfluencies to indicate that they are asking a question or expressing uncertainty. These speech resources are not available to AAC users so it is very challenging for their partner to understand what type of stance the AAC user is taking about a past topic or new one.

O2: Opportunities to Support Mutual Understanding

Led by COM1, COM2, and COM3, workshop participants discussed how they could support AAC and non-AAC users in avoiding misunderstandings and repairing them by keeping track of the conversation topic by making an AC's communicative intent salient to their communication partner:

Identifying topics and communicative intent. Workshop participants mentioned having a way to support ACs and their partners in communicating more clearly by having a way to pinpoint a specific topic an AC is addressing. NLP researchers suggested exploring conversational topic pickers that could enable a participant to speak to a detected topic in the conversation. AC1 agreed to this idea stating that it would be very useful to them by typing: *"100% agree"*. Nonetheless, COM3 and NLP3 pointed out that a topic-driven approach could fail to detect nuances or small topic changes, some topics are easily distinguishable but others are not. COM3 suggested focusing on identifying what an AAC user's stance towards a topic is beyond focusing on words that might fail to represent what the AC wants to convey. COM3 explained that misunderstandings arise not because the topic is unknown but because is hard to link the topic to an intended action:

"sometimes when we know the [communicative] action we do not need to know the topic, what matters is action and people can do action with different words. Sometimes is more important to know what the person was doing than what it was said."- COM3

Augmenting common ground. Participants proposed ways to make an AAC speaking intention clearer to the partner by providing in-the-moment visual contextual cues that can support them. Shared context support strategies mentioned included having a shared visualization of the conversation that could aid partners in making more accurate guesses when trying to solve a misunderstanding. Other ideas involved a topic marker that would visually establish when a current topic wanted to be revisited by an AC. Other ideas included having a one-word reminder of what the topic of conversation was once the AC started typing. Using visual cues to support AAC users in changing topics or facilitating going back to a different one were proposed as ways to reduce misunderstanding by reminding others about details in the conversation an AC may be revisiting.

Ideas like having a visualization of a "mind map" that could link topics and stances could also help communication partners realize what the AAC user was intending to do about a certain topic and would help them have more information when trying to guess:

"Are you giving information, commenting or asking a question? – I think we often try to do different things so the mind map could help [by showing what the AAC user's action is] - "oh you were commenting", my guess sequences were going the wrong way."- COM2

On the AAC's screen, COM2 also suggested adding the ability to easily edit just one word of their message in an easier manner. This editing action could also be sensed by a system and make others aware that the AC's is editing their message and not changing the topic.

Visualization of on-going conversational topics. During Day 2, Team 2 proposed solving the problem of *"contextualizing the AAC user's spoken messages and limiting the interjections of the oral speaker during composition time."* Their idea consisted of having a graphic visualization of ongoing conversational topics in the form of a timeline that displayed the oral speaker's contributions and highlighted the time an AC begins composing a message. That way the oral speaker can know what was being talked about when the AC began typing and it also provides visual feedback to the oral speaker about how many topics or utterances have been produced over the course of the AAC user's turn. By using this tool the team envisioned achieving better mutual in-the-moment understanding. Reflecting on how the technology used for the concept could go wrong, the team identified that some people may feel it is an invasion of their privacy for the conversation to be displayed.

6.3.3 C3: Accessing New Vocabulary

Many AAC devices have a limited set of vocabulary loaded in the device represented via icons or words and organized into category vocabulary pages (e.g., school, home, cooking), but the limited set of words limits what an AAC user can say. If an AAC user wants to use a new word, ACs need to add it to their vocabulary in advance of the conversation, or “talk around” the word in the moment which requires time and effort. However, when a word is not on an AC’s device, they may not recall the word or how to spell it. AC1 described searching for a word online to add it to his vocabulary:

“I have trouble spelling out new words, but I know it in my head. When I am writing something, if It isn’t in my device then I don’t know [how to] spell it then I go to google then write something ”people in love or good friend.” I wanted word to be ”relationship.” then I put in my device. new words is hard for me because I can’t spell and some times I have trouble reading too. my computer read to me and some times if I like a word in there I put in my device.” - AC1

AC1 also gained inspiration for new words to add to his vocabulary by reading and writing down words to add to his device later. Even using a keyboard-based AAC device that does not have a limited word set, AC2 described that it was still difficult to use new vocabulary as his word-prediction tool would suggest only words that he had already used in the past. Beyond AC1 and AC2, one NLP researcher who worked with children who used AAC devices noted that limited support for new vocabulary can also impact children learning to communicate:

“Children often pull random words from their environment and make new words but AAC devices lack explorability.” - NLP1

Even if a vocabulary word is already in an AAC device’s vocabulary, HCI1 and AC2 mentioned that it can be difficult to find if the AC does not use the word frequently. This challenge area was ranked as high priority by both AC participants (see Figure 6.1).

O3: Opportunities to facilitate access to vocabulary on the go

To enable ACs to use descriptive words, participants raised opportunities to make it easier for ACs to add new vocabulary and retrieve existing vocabulary.

Adding new vocabulary. During Day 2, Team 1 suggested and iteratively refined an idea for a word suggestion system where when an AC typed a word, the system would suggest alternative words that may be more expressive, sophisticated, or precise. Team 1 also noted that ACs should be able to turn this new word suggestion mode on or off, and dynamically add words from the word suggestions to their device’s vocabulary.

NLP1 and NLP2 suggested playful approaches to add vocabulary. NLP1 suggested a game for exploring new vocabulary with children where the “AAC user takes a turn developing part of a word and the conversation partner develops the other part”. NLP2 suggested relaxing word prediction approaches to generate more playful responses (e.g., rather than limiting to what the user said, encouraging more diverse suggestions).

Retrieving existing vocabulary. To alleviate cognitive effort when trying to remember where the word is stored in the device, participants suggested integrating search techniques to the device like looking up a word by using phonemes (i.e. “it sounds like this” - NLP3) or a description:

“You could have a word finder if you don’t know where the word is stored you press a button and then you start describing what this word is for and then a bar provides you with that word.” - COM1

AC1 shared that he really enjoyed using Amazon's Alexa to retrieve jokes and control smart appliances in his home. NLP1 and NLP2 build on this to suggest sourcing pop-culture phrases from movies and books as well as important ongoing world events the AC could use in conversation. Other ideas included having the AAC device be aware of keywords AC1 used routinely, similar to how voice-assistants operate, to then retrieve sequences of phrases and vocabulary associated to the AC1's daily routines.

Participants raised that relying on internet connectivity to access a larger online vocabulary or prediction methods to suggest new words may cause the system to not work in low-connectivity settings. For AAC devices, storing prediction models or vocabulary on-device may not work due to limited storage capacity. Finally, participants raised the concern that navigating through excessive or erroneous vocabulary suggestions may slow down the AAC user, as NLP3 described:

“If AAC makes use of text completion algorithms, they may make poor suggestions. Context matters. Not sure but this may actually slow down AAC users who must assess the options while also managing the interface and there might be pressure to accept suggestions that are not exactly what the AAC user really wants to say.”-NLP3

6.3.4 C4: Partner Feedback

Experienced communication partners like speech-language pathologists (SLPs) or friends and family may be able to interpret and attend to specific cues displayed by augmented communicators (e.g., facial gestures and gaze behavior) but unfamiliar partners may struggle to know that an AC is focused typing, trying to get their attention, or listening attentively. Inexperienced partners may lack an understanding of the “culture of AAC”, a term that was often discussed in the workshop to refer to best practices on how to communicate with AAC users (i.e., being patient, waiting for an answer, following up with clarification questions). Unfamiliar conversation partners have few resources to achieve common ground [30] with augmented communicators because they haven't spent time with AAC users to develop this shared background. COM3 explained the challenge: *“ how do we help unfamiliar partners share the common ground [knowledge of each other and conversational norms] plus communication resources (gestures, eye gaze signs) that an AAC user and their familiar partners have in common.?”*

When a communication partner does not notice an AC is typing they may move on, leave the conversation, get distracted or assume the AAC user is not engaged in the conversation. On the other side of the experience, the AC may not be able to see that their partner is confused as they are busy typing on their device. AAC users often do not have physical access to the “grounding movements” produced by the non-AAC partner (nodding, facial expression, etc) while their partner is speaking and they are typing. AAC users might miss what their non-AAC user is saying nonverbally. Participants highlighted how these behaviors lead to a lack of “engaged reciprocity”, being able to show mutual attentiveness that helps build common ground. Communication scientists explained that a lack of “engaged reciprocity” can negatively impact social interaction and cause misunderstandings.

“Usually, when a speaker is saying something, the recipient is looking at them and making backchannel cues, nods, etc. This is sometimes called engaged reciprocity. An AAC user may be engaged in typing, and not have resources for actively showing they are receiving what is being said...There seems to be a communication gap between non-AAC speakers and augmented speakers because there is a technology in the middle. How can the AAC user signal the other and viceversa?”-COM3

Using the metaphor of emojis which are both symbols but also can carry cultural weight, COM1 shared how in AAC there are nonverbal actions that are common like looking at the AAC device when typing but they do not carry the cultural weight that is universally understood. There

is not such a widespread awareness of “the culture of AAC” and this makes it harder for AAC users to interact with new conversation partners.

“we attribute meaning to what we have shared communally... with the waiting is now a cultural expectation. I want you to wait for me because it will help advance communication but also because this is culture, it is what we do. There is a cultural expectation with emojis, but how do we do that for AAC?” - COM1

O4: Opportunities to support multi-modal partner feedback

We considered how language tools could serve the purpose of mutual signaling and back-channeling more salient to all communicators. Participants brainstormed different ways to increase engaged reciprocity by using “thought bubbles” to signal typing and visual alerts or flags to signal “hold the floor I have something to say”, or auditory feedback to alert a communication partner. These flags could be automated to detect typing events or specific words in order to not place an extra burden on users. COM1 shared how some AAC users already use signaling to bring attention to their communication:

“I remember being in a school where a child only had access to a low-tech board. And he would point to a symbol with his left hand while using a plastic toy to bang the table with his right. I thought it was so genius, but was largely ignored by teachers as disruptive, but it was so clear he only banged the toy when also pointing at symbols. Would be interesting to think of other ”I’m composing/hold the floor” alerts/flags/etc.” - COM1

Multimodal tools to cue conversation partners. Workshop participants shared ideas around tools to cue conversation partners including monitoring the conversation turns and using this information to remind partners to pause either through programmed interactive reminders or a third-party agent, a “computer moderator” that could act as a referee or mediator between non-AAC and AAC participants and would be in the “look-out” for cues a partner misses.

6.3.5 C5: Managing Turns and Timely Participation

ACs need more time than oral speakers to participate in conversation. They first type and then play their message. In group conversations, there is often a social expectation for fast speaking so an AAC user might feel pressured to give a short response fast. AAC users get to say less than oral speakers, and the lack of social support leads them not to feel like people care or are valued:

“I need more time to type my questions and answers. Some people just don’t care what I, or whomever, have to say.” - AC2

Staying in time is a well-known and still unresolved challenge for augmented communicators. ACs often fall behind in a conversation as other speakers are able to participate at higher speeds and without long enough pauses leaving augmented communicators behind and reducing their chances to participate [157]. When conversations move on quickly and an AC fails to take the turn at the appropriate point in time they may be misinterpreted by speaking “out-of-context” [125]. AC2 stated “*I am always a conversation behind.*”

AC may also hesitate sharing their message out-loud because they might interrupt the flow of the ongoing conversation. Often AAC users typed in the chat during the workshop to avoid interrupting while still voicing their ideas.

O5: Opportunities to augment turn-taking with interactive tools

Participants brainstormed ideas that could support AAC users in replying in time but also their conversation partners in waiting patiently. Participants discussed having user interfaces that adapt to the current situation to support fast retrieval of relevant responses and monitoring and visualizing all speaking turns to help partners remember to wait.

Visualizing different phases of a turn. Technologists and communication scientists suggested that visualizing speaker turns and user typing activity could help partners pace themselves and be reminded to wait. For example having a shared visual display that can track context and help everyone know that there might be some people falling behind in conversation might slow the non-AAC partner (ideas shared by NLP2, COM2). Pointing out where ACs start composing within a conversation, would make conversation partners slow down if they are also aware that they might be interrupting the composition. Technologists further suggested needing to identify discrete sequences of conversation, have labels for each speaker, and additional labeling for each type of participation event that wants to be visualized. One great limitation to these approaches was a concern as to how and where all of this information was going to be displayed.

Conversation-aware and adaptive AAC. AC1 suggested having the AAC device's interface dynamically changed based on questions he gets asked. AC1 envisions that his own AAC user interface will adapt to the current conversation and update itself to support him in easily retrieving the most relevant responses suitable for the situation. AC1 explained:

“if you ask AAC user a question. Hey, how's it goin'? a device change display screen itself. Change to feels area, but what happens if a person AAC user in the workshop or event then a device is going crazy? I think that it will work on one on one.”

Technologists agreed that a system that learns to recognize specific questions or phrases could automatically update the interface to show candidate options. Additionally, technologists suggested other ideas like having the interface follow-up to a previous conversation:

The device is able to 'remember' the last conversation that the AAC user had with a particular partner to help them continue the conversation from where they left off. For example, the partner mentions that they have a job interview coming up the next day, the next time they talk the AAC user may want to ask how the interview went. - NLP1

Both of these ideas could support new ways of accessing and retrieving pre-stored vocabulary and messages.

6.4 Discussion: a Research Agenda for NLP in AAC

Communication scientists and ACs highlighted that there is an opportunity to promote AC communication through social support of communication partners without further technology (e.g., noticing typing and waiting for a response). To complement improved social support, we outline opportunities for future research that aims to further participation in conversation for augmented communicators:

6.4.1 Using Conversational Context to Reduce Misunderstandings

Traditional approaches for ACs word prediction rely on only the previously typed words by the AC [64], or a physical location [78], to predict potential next words and phrases. Our workshop revealed a potential opportunity to help ACs compose their messages based on the prior conversation

to improve the specificity of short, out-of-context messages that are ambiguous to conversational partners who do not wait. While prior work has considered using conversational context to improve AC's speed (*e.g.* by extracting keywords from the conversation to generate sentences from AC's sentence acronyms [27]), our workshop points to additional opportunities to use conversational context to reduce misunderstandings by improving: (1) the technology to enable ACs to add specificity to their responses, and (2) the social support for ACs by helping conversational partners recall the context of ACs speech.

Transcribing the Conversation Speech.

While transcribing speech to text could enable generating responses based on the conversational context [27, 166], workshop participants highlighted that keeping a display of the prior conversation transcript itself (*e.g.* on an AAC display or phone app) could help conversational partners repair misunderstandings by referring back to the prior conversation. In addition, pausing the transcript where the AAC user started talking could help partners understand the context of the conversation in which the AAC started composing their response (*e.g.*, the conversation has moved to talk about dinner but the display shows that the AAC is responding to an earlier question “how was your morning?”).

While prior work used conversational context from text [27] or conversation partner utterances in a controlled lab setting [166] or through text input by the partner on a companion app [46], transcribing speech in real-world environments will introduce errors (a concern for workshop participants) as transcription accuracy depends on the speech itself (*e.g.*, common vs. expert words), microphone placement, and environmental noise [123]. Because transcription errors can propagate to other steps, using conversation context may be most readily applicable for text-only chats (avoid transcription) and virtual meetings (fixed microphones). Research to improve speech-to-text transcription should include AAC voices in their datasets to enable such future applications.

Conversation Keywords and Phrases.

To automatically generate relevant responses to respond to text conversations, prior work in dialog research has detected keywords [138] or dialog acts [115] (*e.g.*, asking a question) in the prior conversation context to guide response generation. Cai et al. similarly used text keywords and AAC user-generated sentence acronyms to guide response generation for AC keyboard users. However, our workshop indicted alternative uses of such technology to reduce misunderstanding including: using extracted keywords from the conversation to dynamically generate word pages with words likely to be relevant to the conversation (similar to the static category word pages used by many ACs today), or using extracted keywords to display to the conversation partners what keywords appeared close to when the AC started their response (*e.g.*, “AC1 is talking about dinner”). These ideas could connect to work that seeks to support mutual understanding for people who are deaf or hard of hearing using word clouds displayed to all participants during remote meetings [71].

Rewriting Ambiguous References.

ACs short phrases expressed after the conversation has changed topics (*e.g.*, “she went there”) can contain references (*e.g.*, “she”, “there”) ambiguous to conversational partners. Co-reference resolution [88] or event correspondence [114] could be used to detect the likely original subject of ambiguous references (*e.g.*, “Mari”, “Canada”) then display the results to partners or the AC who may choose the word or a sentence with the reference replaced (*e.g.*, “Mari went there”).

6.4.2 Controlling Text Generation to Improve Expressivity

Beyond generating text to improve speed [27], our workshop revealed that ACs would like to generate more expressive speech with accurate grammar and pronunciation. Prior work has explored adding emotion to AAC utterance through emoticon-based keyboard notation [45]. Given the increasing availability of large language models that can learn from patterns, a model could also learn from the spelling strategies AAC users use to bypass common text-to-speech engine limitations.

Identifying word alternatives.

Given ACs enthusiasm about speaking more precisely, commercial AC devices can explore low-compute approaches to improving expressivity, by embedding thesauruses to look up synonyms for words with a “word suggesting mode” as suggested by workshop participants. Such approaches could be refined by considering the probabilities of AACs using different words given the conversation and their text so far (*e.g.*, “fine” may be a more likely synonym for good than “upstanding” if the AC is discussing breakfast).

Creative text generation.

Word and utterance predictions are currently used in AAC but sometimes are limited to current vocabulary already known and commonly used to the user. Advances in large language models enable an interesting opportunity of balancing the advantages of a fixed vocabulary with the possibilities of playful expansive vocabulary and phrasing by incorporating input from a model at different moments of the message composition process [169]. Creative writing support via algorithms has also been explored in the NLP community [53]. Text generation approaches for AAC users’ could consider expansive rather than safe predictions to enable more expressive speech.

Controlled text generation.

Alternatively, future work may explore how to enable ACs to say their statements in more expressive or grammatically correct ways by controlling response generation. For example, future work may generate alternatives to ACs’ current message based on emotions [171], style [80], or preferred politeness [108] by extracting message keywords and using an additional input (*e.g.*, formal) to rewrite a message to be appropriate for the context. As suggested by workshop participants, the concept of “communicative stance” from communication science suggests a promising mechanism for control (*e.g.*, “disagree”, “agree”). Such rewriting can also be used to improve grammar of responses if the AC desires (*e.g.*, similar to Grammarly).

Speech generation.

In the last decade text-to-speech pronunciation has improved by moving from complex pipelines with many steps (*e.g.*, words to phonemes, phonemes to pitch contours) for statistical parametric synthesis [148] to end-to-end neural speech synthesis approaches that pronounce speech more reliably [85, 117]. Thus, updating text-to-speech approaches will likely improve ACs experience with pronunciation in addition to low- or no- compute approaches like crowdsourcing a dictionary of pronunciation (*e.g.*, similar to YouTube pronunciation guides) and prompting partners to accept mispronunciations (*e.g.*, hold laughter so that the AC can finish speaking).

The workshop indicated ACs desire for greater expressivity, and prior art by Graham Pullin titled 17 ways of saying ‘Yes’ [119] indicated an opportunity for furthering expressivity with pronunciation. AAC devices could apply prior research on how to adjust the tone of generated speech using low-level approaches such as SSML [149], a markup language to manually specify low-level characteristics of speech playback such as pitch and speed (which may require too much

input). Alternatively, high-level approaches such as style-controlled speech generation that can apply the style of an example speech recording to a new message [163], could enable ACs to store styles they would want to use (*e.g.*, an upbeat or formal example) then apply the styles to their messages to adjust pronunciation. However, such generation may be limited to labeled examples that the AC saves in advance. In the future, developments in prompt-driven audio generation (following a pattern of developments in prompt-driven image generation and prompt-driven text generation) could enable AC users to creatively apply new speech styles on the fly.

6.4.3 Limitations

One limitation of our workshop is that we did not spend time considering a broad perspective of possible concerns users may have about sharing text logs of all their communication in order to make many of the ideas suggested possible. When talking about proactive technologies in our workshop we shared with AAC user participants that these technologies would have access to their data. Nonetheless, we did not have a focused discussion on AAC users' perceptions and level of comfort with sharing their communication data. Future work should continue to carry out conversations with stakeholders about possible social consequences of integrating language technology solutions in AAC users' lives. We acknowledge that many of the solutions proposed need continuous discussion and continuous re-examination [145].

6.4.4 Conclusion and Future Research

Our workshop started with ACs sharing about themselves and their communication. Their stories continued to shape the workshop discussions and helped ground our explorations in challenges they identified as most pressing and solutions that would bring value to them. Communication scientists, many with backgrounds in speech and language therapy brought their years of experience supporting many AAC users achieve their communication goals. Additionally, communication science participants enriched our discussions by sharing their technical knowledge (*e.g.*, theory around repair, grounding, communicative stances) that allowed us to better articulate causes and nuances in communication breakdowns. To complement, NLP researchers provided rich accounts about current state-of-the-art language technologies, suggesting ideas and identifying opportunities. HCI researchers proposed the structure of the workshop, and carefully designed the activities to elicit responses from all participants. All participant types interacted with each other as a large group and within interdisciplinary teams to discuss the problem space and what was feasible and practical to achieve. Our participatory exploration suggests that social communication goals such as how speaker and addressee are able to stay in time, manage turn-taking, and achieve mutual understanding are as important as developing efficient and usable AAC systems. The set of experts that we invited was chosen given our focus on language technology and AAC. Future iterations of this work could include other disciplines to explore different intersections of the problem space with other fields (art, interaction design). All in all, bringing all of these disciplines together proved beneficial as a generative design strategy. Our work demonstrates great value in using design strategies to scaffold conversations between different disciplines and experts to generate new research directions.

6.5 Summary

Our two formative studies indicated possible directions to explore to support mutual understanding and participation in conversation. First, the time at which an AAC user starts typing can be a strong signal that can help find the potential relevant context portion in the conversation. Due to different AAC typing strategies to stay on time (*i.e.*, typing while others speak, or typing in chunks) a tool that supports them in group conversation has to take into account the ongoing conversation

but differentiate the instances when the AAC user is typing. Similarly, if a tool has access to the current conversation perhaps it can leverage this information to suggest possible references to different topics in the conversation. If the tool is able to create a share visualization of the conversation, perhaps this can support grounding between the CP and AAC user and the non-AAC communication partners can more easily pace themselves in conversation. Additionally, enabling the tool must have a feature that allows the AAC user to specify their stance or intention towards a topic or theme in the conversation, as this would make possible reference suggestions more accurate and could help create better notifications or feedback cues that are more specific to the partner. Having access to this suggestions may allow the AAC user to participate more often in conversation.

Chapter 7

How Large Language Models can Support and Hinder Communication

7.1 Overview

This work continues the exploration of the previous chapter, looking into how Large Language models can, if at all, support AAC users in addressing some of the challenges identified with regards to message presentation and message composition.

A primary strategy for improving AAC device performance is to predict what the user intends to type and offer it as a suggestion [154]. These predictions can come from many sources, including static language models [154], photographs [48, 47], or contextual information about the user [76, 78]. AAC users themselves may attempt to predict what they will discuss in the future and pre-write messages that they can later retrieve via their AAC device [77].

Recently, advances in large, neural language models (LLMs) such as GPT-3 [24] and BERT [39] have created new opportunities for improving the usability and efficiency of AAC devices. Current LLMs are able to generate text that is indistinguishable from text written by a human [29], potentially enabling AAC users to generate human-level speech with minimal effort. Preliminary research with simulated user data has shown that LLMs can retrieve contextually relevant sentences [138] and expand user abbreviations [27], theoretically reducing an AAC user's keystrokes by up to 75%.

While these potential gains are encouraging (and in fact will likely continue to improve), it is important that AAC users be involved in the process of combining LLMs with AAC devices. This involvement helps ensure that LLM output meets the users' expectations, and that interactions between LLMs and AAC users support the users' communication preferences, all while maintaining privacy, autonomy, and control.

In this chapter, we present a study with 12 adult AAC users in which participants generated speech suggestions from an LLM and provided feedback about those suggestions. Our participants had a variety of disabilities that affected their speech production but not their language use or understanding. To support participants' experimentation with the LLM, we introduce the concept of *speech macros*: LLM prompts that transform abbreviated user input into full sentences, with a focus on achieving conversational goals such as requesting help with something or answering a biographical question. Participants tested each of these macros over a remote video call, trying various inputs and commenting on the outputs, and later provided feedback about their experience via an online questionnaire. Our study was guided by the following research questions:

- *RQ1*. What are the benefits for AAC users, if any, of directly interacting with large language models (LLMs)?
- *RQ2*. How do AAC users evaluate communication suggestions made by an existing LLM?

- *RQ3*. What concerns do AAC users have about integrating LLMs into their own AAC devices?

Overall, our participants were excited about the possibility of using AI-generated suggestions in their AAC device, but articulated some requirements for these suggestions to be usable. Specifically they requested that these suggestions are contextually appropriate, match the user’s personal conversation style, and provide the ability to customize, edit, and remove suggestions.

This project makes several contributions toward the goal of enabling AAC users to benefit from the capabilities of LLMs. First, we present a study in which AAC users interacted with an LLM in real time and provided feedback about the suitability of the LLM’s output for their own communication needs. Second, we introduce the concept of *speech macros* as a way to leverage the generative capabilities of LLMs to support the specific communication needs of AAC users. Finally, we identify opportunities and challenges to creating AI-powered AAC systems.

7.2 Integrating LLMs into AAC with Speech Macros

In considering how LLMs can provide useful suggestions to AAC users, we landed upon the concept of *speech macros* as a way to explore scenarios in which LLMs generate content for AAC users. Our approach is similar to KwickChat’s *bag-of-keywords* model [138], in that an AAC user provides one or few input words that are then converted into a complete sentence. However, in contrast to prior work, our speech macros go beyond sentence expansion to support a variety of connections between input and output.

7.2.1 Design Process

We began this project with an exploration of how LLMs could support use cases common to AAC users, and how we might explore those benefits in the context of a user study. We conducted several brainstorming and sketching sessions within our research team, which contains HCI/accessibility researchers, researchers with experience related to LLMs, and speech language pathologists.

Through this process, we identified a set of potential benefits that LLMs can provide to AAC users, including some that have been explored in prior work:

1. ability to create full sentences from abbreviated input (as explored in [138, 27]);
2. ability to draw from conversational or user context (also explored in [138, 27]);
3. ability to generate grammatically correct sentences in response to a question (an identified need mentioned in Chapter 6, section 6.2);
4. ability to customize the tone and content of output.

While our prototype includes elements of all of these benefits, we ultimately decided to focus on how LLMs can be instructed to perform a variety of tasks using natural language prompts. For example, prompts provided to a general purpose LLM can be used to quickly prototype French-to-English translation [74]. Prior research about AAC users has often identified challenges with specific forms of communication, such as when talking to a physician or telling a long story [78], and that AAC users often conduct extensive work before a meeting to prepare what they wish to communicate [77]. Thus, we chose to explore how specific conversational tasks, such as requesting help with a particular object, or answering questions about one’s background, could be supported by prompting an LLM. This approach is complementary to work that is focused on improving AAC expansions in everyday conversation [138, 27].

7.2.2 Speech Macros

We created Speech Macros to act as boundary objects and design probes to exemplify LLMs' capabilities and to demonstrate real-time output based on different conversational situations and user inputs.

Speech Macros were designed to be purpose-driven shortcuts that can generate complete sentences from a brief input, such as a single word. Informed by prior work that uncovered challenges in AAC-based social interactions [158, 65, 77], we created multiple prompts using the transformer-based large language model LaMDA developed for dialogue applications [151]. This model's output can be customized through zero and few-shot prompts. In our tests, we found we could provide 1-3 examples and a description of the desired output to produce reasonable results (see the Appendix for the prompts we used). The model produced a variable number of responses, which varied in length from a few words to several sentences. However, since the multiple sentences generated by the model often contained unrelated "hallucinations", we delimited each response to include only the first sentence generated. We restricted our macros to showing the first four responses generated by the model so that they fit on the screen without requiring the user to scroll, and to provide a manageable number of suggestions to read and evaluate, comparable to the number of suggestions provided by existing AAC systems.

After testing different prompts through word choice and example iterations, we generated three Speech Macros that produced phrase suggestions for users based on different available contexts, underlying task instructions, and user inputs. We selected Speech Macros that performed well under different conversational situations, and with different types of user inputs. We then created a web-based prototype for each macro (Figure 7.1).

Extend Reply

Phrases produced by LLMs can leverage specific conversational context, like the ongoing dialogue, to provide specific responses that can help reduce misunderstandings among conversation partners while helping the AAC user be more specific about what they want to say. Motivated by the known problem that current AAC input methods may limit how detailed an AAC user's response can be (as more detail means more effort), the first Speech Macro, *Extend Reply*, extends a user's short input with more details that fit an ongoing conversation. To demonstrate this LLM functionality and to support users in sharing more detailed responses with less effort, the Extend Reply prototype has three main features: (1) a place where we represented the model knew what a conversation partner had just said (the current conversational context), (2) a place for user input to respond to the current conversation, and (3) suggested phrases by the model generated based on the instruction to extend the user input into a contextually relevant sentence that could be used in conversation.

Reply with Background Information

In addition to supporting user input during a conversation, we explored the possibility of allowing users to fill out information ahead of time and use that stored information to generate suggestions in a later conversation. During the study, we asked AAC users how an LLM could reduce their effort, and several participants mentioned that they often get asked the same questions repeatedly throughout the day, and prior work has shown that AAC users often write out things that they might want to say before a meeting so that it can be quickly retrieved during the conversation [77]. The *Reply with Background Information* Speech Macro accepts a paragraph of text in which the user includes information that they might wish to retrieve later. When they are asked a question, the AAC user can generate responses based on the previously supplied information. To maintain our interaction model of combining a conversation partner's question with user input to generate a response, the user does not enter additional text in this example, but instead presses the button

to automatically generate potential responses, although future versions could certainly combine stored content with live input.

One feature of this Macro is that the system can automatically generate responses that match the phrasing of a specific question, regardless of how they originally wrote the information. For example, an AAC user might include a declarative statement in their bio such as “I have a cat named Kevin.” If the conversation partner asks “Do you have any pets?” the system would reply with “Yes, I have a cat named Kevin.”, while if the conversation partner asked “What is your cat’s name?” the system would respond with “My cat’s name is Kevin.”

Turn Words into Requests

Another important key functionality of LLMs is that they can be prompted to complete specific tasks such as turning a word into a help request. We wanted to communicate this functionality to AAC users so we developed the *Turn Words into Requests* Speech Macro. We imagine AAC users could create their own instructions or prompts in the future to retrieve outputs from a model that fit their needs. The Turn Words into Requests Macro prototype consisted of only two components: (1) a place where the user could input a word they wanted to ask help with and (2) a space to see generated help request suggestions.

7.2.3 Prototype User Interface

Even if an LLM can be trained to provide high quality suggestions, there remains the challenge of integrating LLM feedback into the AAC user interface. At the same time, conducting early stage design studies requires communicating a lot of information to the user under significant time constraints.

For this study, we chose to sidestep any detailed questions about the user interface, and instead focus on a prototype that enabled AAC users to test the model and evaluate its output. Our prototype features a minimal user interface that highlights three main components: a question from a conversation partner, user input, and suggested phrases from the LLM (Figure 7.1). The user is able to change either of the inputs and regenerate the suggestions; thus they can explore how different inputs lead to different suggestions, or how a particular input would function in response to different questions. For our study, the input fields were pre-populated with example text so that participants could immediately test the system and see live output from the model. Additionally, our prototypes included a *variability* slider that helped modify the model’s output during the study in cases where the model produced the same text suggestions repeatedly. A higher variability value creates more random output. We explained the variability value to users and set it to an approximate mid-point of 0.6 and only changed it when the model did not suggest sufficient phrases or suggested repeated phrases. Our choice of 0.6 was based on multiple testing of our prompts and variability combinations that produced varied phrase suggestions.

7.3 Evaluation Study

Knowing that LLMs can generate diverse outputs from one short set of inputs, we wanted to understand what type of prompts would be most useful to AAC users, what types of inputs they could provide, and what types of outputs were the most useful to support their communication. We designed a user study that would first introduce participants to LLM capabilities, focusing on three main abilities we thought could be the most relevant for AAC users: how a model can (1) suggest words based on conversational context, (2) draw from general world knowledge and (3) learn from examples and specific instructions (prompts). During the user study, participants also experienced the three speech macro design concepts and provided feedback during the study and in a post-study survey.

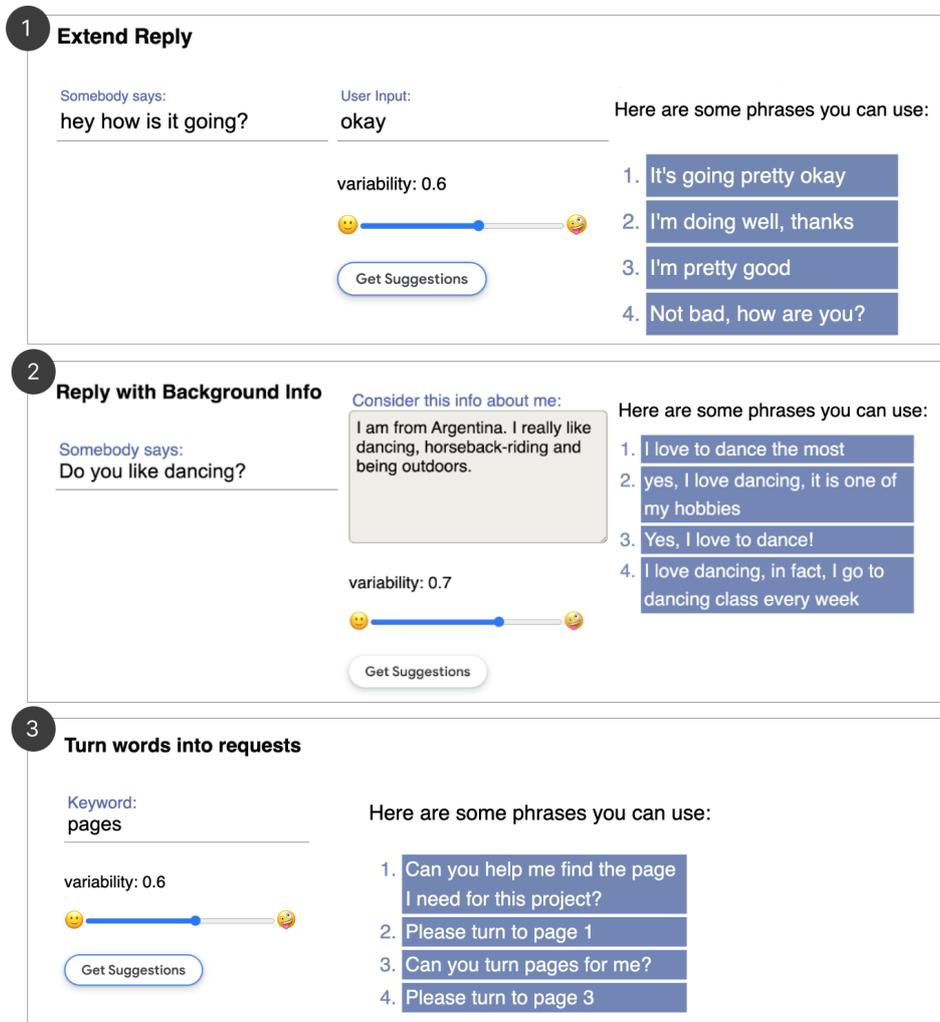


Figure 7.1: Prototype of *speech macros* used in the study. Each screen includes the name of the current macro, the conversation partner’s question context, space for the user’s input, and a ‘variability’ slider to adjust the temperature of the model, and thus the diversity of the responses generated. Output from the LLM is presented on the right side of the screen.

Table 7.1: Our 12 study participants (3 females, 2 non-binary (NB)) used a diverse set of AAC devices and techniques and had different levels of speech use.

ID	Age group	Gender	AAC Device	Years using AAC	Access Method	Non-verbal?
P1	45-54	M	Tobii Dynavox	1	Eye gaze	Fully
P2	45-54	M	Tobii Dynavox	4	Eye gaze	Often
P3	65-74	F	Proloquo4Text App on iPad	9+	Mouse, joystick	Fully
P4	65-74	M	Grid 3	3	Keyboard	Fully
P5	45-54	M	iPad with Predictable App	16	Touch	Often
P6	25-34	F	Type on phone and Google docs	8	Head movement, switch	Sometimes
P7	25-34	NB	Dynavox Maestro	23	Switch	Fully
P8	35-44	F	iPhone and Android phone	5+	Touch	Sometimes
P9	25-34	M	Tobii Dynavox I-15 series	32	Switch	Fully
P10	55-64	M	Speech Assistant App on Galaxy S20+	9	Switch	Fully
P11	25-34	NB	Android tablet with Predictable, Coughdrop, and Speech Assistant Apps	8+	Touch	Often
P12	45-54	M	iPad Pro with Proloquo4Text App	40+	Toes	Often

7.3.1 Participants

We recruited 12 adult expert augmented communicators, who use a variety of AAC devices, to test all three Speech Macro concepts. Our 12 participants (Table 7.1) included two eye gaze AAC users, four switch users and six AAC users who used direct selection to interact with their communication devices. Our participants used AAC solutions for multiple reasons including degenerative chronic illnesses, apraxia, cerebral palsy, autism, and also a combination of all these factors. None of our participants had aphasia or any disabilities affecting language use, only verbal speech production. Participants resided in the United States, the United Kingdom, or Canada.

In preparation for the study session we asked participants if they wanted to join the call with a person that could support their participation. We also asked about preferred communication styles and broadly about how we could make the study accessible specifically to them. Some participants joined the study session with a support person that helped them connect to the video call or communicate. Support persons were often family members or speech and language therapists. We did not consider support persons to be active study participants, but in some cases they did provide comments during the study, and we include those in this chapter when appropriate.

7.3.2 Procedure

We gathered participants' feedback through a 90 minute remote video call and a post-study survey. We divided the remote study session in three main parts: (1) introducing language technologies, what they are and how they appear in some products (10 minutes), (2) testing the three speech

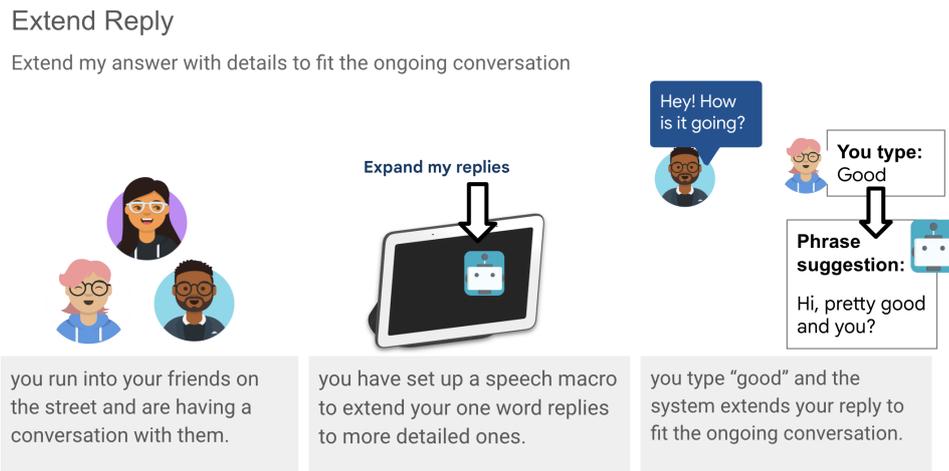


Figure 7.2: Storyboard exemplifying a sample interaction with the Extend Reply Speech Macro.

macros to evaluate their usefulness and their outputs (60 minutes), and (3) reflecting on other uses for speech macros (10 minutes). We also offered participants an optional 5 minute break (or more time if needed) that could be taken at any point during the study.

Introduction to LLMs

To introduce participants to LLMs, we presented different examples of AI-based language technologies that use LLMs, such as word prediction and word completion keyboards, auto-complete, and translation software. We explained that the LLM they would interact with during the study had learned patterns about vast amounts of text-based data from the internet. We noted that this made it a useful tool to support conversations since it could “understand” dialogue and suggest possible responses given a specific task or instruction. We introduced participants to Speech Macros as a way to define tasks the model could perform to assist in communication.

Trying out the Speech Macros

We then introduced participants to each Speech Macro by first illustrating a scenario using a Storyboard to illustrate specific use cases (Figure 7.2) and then shared the interactive prototypes via screen-sharing. All participants were introduced to the three Speech Macros in the same order: *Extend Reply*, *Reply with Background Information*, and *Turn Words into Requests*. We first explained each prototype and how these were just examples to test the model’s functionality, rather than finalized communication device concepts.

For the *Extend Reply Speech Macro*, participants were presented with two example conversation scenarios. For the first, we said to imagine a conversation partner asked: “What did you do today?” and they, the AAC user, used a short word (*i.e.*, “work”) to get extended replies from the system. For the second example, we explained that the system could adapt to different situations, such as the conversation partner asking, “Do you want to get pizza?” and then providing the same user input of “work”. After each example, we asked participants what they thought about the suggestions, and whether they would accept any of the suggested phrases¹. After reviewing

¹We originally asked participants to rate each output on a scale. However, participants were frequently unable to choose a rating for a single phrase set, so we omit these individual ratings from analysis and instead focus on the comments they provided after each phrase set, as well as the ratings and comments about each speech macro.

the examples, we asked participants if they would like to try a user input to reply to either conversational scenario. If time allowed, participants could try more than one user input to reply to a conversation situation or to suggest a question they often get asked. At the end of the macro, participants were asked to rate how useful the functionality of having the system extend their reply would be to them and to comment on their rating using a scale ranging from “Not at all useful” to “Extremely useful”.

For the *Reply with Background Information Speech Macro*, participants were first presented with a fictional sample biography: “I am from Argentina. I really like dancing, horseback-riding and being outdoors. I do not like insects. I love to eat ceviche, arepas, and tacos. I have a cat named Stella”. The test conversational scenario was the question: “Do you like animals?” After discussing how the macro worked and rating the output for the example scenario, participants were invited to add their own biographical details to the existing text so that the model could use their own background information. Participants shared information about their favorite animals, favorite sports teams, hobbies, or country of origin.

For *Turn words into Requests*, participants were directly asked to think about any items or actions they would like to ask help with and suggest them to try as user inputs. After trying each suggested input, participants were asked to rate the generated phrase suggestions. After trying various inputs participants were asked to rate how useful was the Turn words into Requests functionality by using a scale ranging from “Not at all useful” to “Extremely useful”.

Post-study survey

A post-study survey was sent to participants to capture any additional open-ended feedback they did not get a chance to share during the study session. The post-study survey contained both multiple choice and open questions and was organized in sections: (1) feedback about specific Speech Macros and ideas for additional speech macro functionalities not covered, (2) feedback about the concept of Speech Macros (where they would be useful and where they would not be, what were some benefits, what were some concerns), (3) Priorities for future versions of speech Macros (important and less important use cases); (4) Using personal data in AAC (concerns about personal data use and information they would feel comfortable sharing), and (5) General feedback (Any other things you wish an AI-based communication system could do for you? Any additional feedback you would like to share?).

7.3.3 Data Collection and Analysis

Study sessions were video and audio recorded; audio recordings were transcribed using automatic speech recognition and corrected manually by the research team. Transcripts of the session were combined with a log of the session’s text chat and researcher notes into a single document. Participant responses to the post-study survey were stored in a separate document.

Two members of the research team analyzed the data; both have several years experience in human-computer interaction and accessibility research. Both researchers had prior experience in conducting participatory design research with AAC users. One researcher had experience using accessible technology in their everyday life.

We performed qualitative data analysis to organize the findings and identify common themes [22]. First, the two researchers independently read through the 12 transcripts and post-study surveys, selecting quotes and observations and copying them onto separate notes, which were organized through several rounds of affinity diagramming [103]. We identified four categories of data: feature suggestions, potential use cases, comments about the quality of suggested phrases, and observations about using AI-enabled AAC in daily life. Feature suggestions and use cases were organized by which macro they related to, and are mostly presented in Sections 7.4.1 and 7.4.2. The remaining data were analyzed through several iterations of discussion, note-taking, and affinity

diagramming, beginning with identifying the most common high-level themes in the data (characteristics of good/bad sentence suggestions, AAC use as self-expression, how AAC influences perceptions of its users, and concerns about AI) and grouping data into subcategories within them. These themes are largely discussed in Sections 7.4.3 and 7.4.4.

7.4 Findings

We first report on participants' experience using Speech Macros, the inputs they tried for each macro, and the suggestions participants provided. We then report on participants' feedback about the model's output, and lastly on key user concerns to consider when integrating AI-based language technologies into communication devices.

As our participants had different speaking rates, we tried to spend an equal amount of time discussing each Speech Macro with each participant (around 20 minutes per Speech Macro). Some participants with faster speaking rates provided more user inputs of their own, while others provided at least one user input or conversational scenario suggestion. The statistical Median of interactions with the *Extend Reply* macro was three, and two for the *Reply with Background Information* and *Turn Words into Requests* macro.

All participants completed the post-study survey and provided extended written replies and thoughts (Median: 376.5 words, Max: 1023 words, Min: 43 words). Only three responses were less than 100 words while all the others were above 250 words.

7.4.1 Benefits and Uses of Speech Macros

Participants found the conversational tasks that the speech macros supported to be in general very useful to them (Figure 7.3). From all three speech macros, the "Turn words into Requests" was rated more often to be either extremely or very useful followed by "Reply with background information" and then "Extend Reply."

Extend Reply

Participants liked only having to input a few words to get phrases extended by a macro since it could help reduce typing effort and fatigue. *"[the extend reply macro] would enable me socialize and network more because I would be able to type faster and would require less time effort and energy and also lessen frustration, sometimes I just don't initiate conversations because I don't have the energy to type a lot and I can't answer quickly enough"* (P6). P11 also shared that phrase suggestions could help them alleviate cognitive effort in trying to remember what to say that could appropriately fit the social situation. *"That sounds super helpful for knowing what words to use that are socially acceptable rather than getting stuck and trying to remember and at the same time trying to go through the physical actions of using AAC."*

The Extend Reply macro generated alternative responses to the same question, for example "I'm hungry" or "I'm not hungry", to reply to "do you want to get pizza?" (Table 7.2). P6 really liked having options as it would allow her to choose different suggestions depending on her mood, the circumstance, and the person she is addressing.

Given that macros were created using few-shot prompts that were straightforward expansions of an input (see Appendix for prompts), some phrase suggestions seemed too cold-cut for social conversation. When asked to rate how appropriate the model's suggestions were, participants brought up needing more information about the scenario: would they be using the extend macro on a mobile phone or a computer? (P5) who are they addressing?(P2,P3,P6,P10) what is their mood?(P6,P11). For scenarios in which they imagined talking to their friends, participants preferred phrase suggestions that were casual and friendly instead of more straightforward options (Table 7.2).

Table 7.2: Extend Reply Speech Macro inputs given by our study participants, the suggestions returned by the speech macro, their preferred option (shown in bold), and their rating of how appropriate that suggestion would be for the question asked and their personal communication style.

Question	ID	Input	Suggestions	Rating
Do you want to get pizza?	P6	Later	<ol style="list-style-type: none"> 1. I'm feeling pizza later, are you? 2. What time do you want to get pizza? 3. Sure, let's get pizza 4. Ok, where do you want to go? 	Extremely appropriate
	P1	Eat	<ol style="list-style-type: none"> 1. I'm always up for pizza 2. I'm sorry, I'm not hungry right now 3. I'm not hungry 4. I'm hungry 	Very appropriate
	P3	Work	<ol style="list-style-type: none"> 1. I'm sorry, I can't go out this afternoon 2. I'm already at work 3. I'm busy with work today 4. I'm working today 	Not at all appropriate

Reply with Background Information

One strength of LLMs is that they can draw general information about the world into their generated responses. This ability seemed to be very useful when the participants tested how the model suggested phrases for them in the “Reply with Background” macro. The ability to guess and imply information from a short biography provided by participants seemed helpful when it related to factual knowledge about the world. For example, when telling the model that she was from Sri Lanka originally, the model made suggestions like “I love to go to the beach” when replying to the question “what do you like to do for fun?”

“those activities are things that I would do back home. It knew somehow that because I am from Sri Lanka, that is something important to me, so I love that it took that piece of information and took it as my background. I think that is important.” - P8

In another case, P9 mentioned his favorite soccer team, the *Chivas*, and was pleased when the system was able to identify the location of the team, as it suggested the phrase “my favorite team is Chivas de Guadalajara”. While P9 was impressed by the system’s knowledge, he ultimately chose a different response that better conveyed his lifelong enthusiasm for the team: “*My favorite soccer team is Chivas, I have supported them since I was ten*”.

On the other hand, the language model often guessed and made up information to fill in the details that seemed random and unrealistic or “wrong funny” as expressed by P11. The model guessed that P11’s dog was “a cute little dog” based on its name “Pippin,” which P11 provided in their background information. In reality, Pippin is a 100 pound German Shepherd. “*The false assumptions would make me want to go add bio information later but not right then.*” - P11.

Participants shared examples of meaningful context they would want the system to draw from to make more useful suggestions. Participants wanted the model to know medical information about them so that they could have informed conversations with their medical doctors (P10, P11). Participants also wanted the model to automatically draw information about the time of day and their location (P7, P10, P12). Some AAC users wanted the device to know about their favorite topic, like their favorite soccer teams (P5 and P9), and even be aware about common topics shared with specific conversation partners (e.g., co-workers, friends):

“I think background information could encompass a great deal beyond demographics. For example, being in IT includes a kind of sub-vocabulary relevant to speaking with colleagues, as a chess player there exists a sub-vocabulary, and the roles of husband, father, grandfather, church brother, and more could each have identifiable “background information” that could influence phrase generation.” - P10

Turn Words into Requests

The “Turn Words into Requests” macro worked well when requesting help with inputs about common tasks (i.e., tea and biscuits, bathroom, sleep), and was described positively by participants indicating that it could be useful to them (Figure 7.3). Nonetheless, when receiving input like “smoke”, in a case where the user wanted to try to request his caregiver to take him out for a smoke, the model steered the user from this specific activity. During the study, we tried different input including smoke and cigarette and for both tries the model suggested phrases such as: “no smoking in the house”, “I do not want to smoke”, “please don’t smoke a cigarette”. We speculate this subject was caught by a safety or policy layer intended to avoid promotion of this specific activity.

We found the system was not able to suggest relevant requests related to access or health needs. For example, when P3 tried the word “transfer” to see if the model could suggest requests related to needing help transferring from her chair to the bed using her home-installed lift, the model only suggested phrases related to transferring money (i.e., “transfer money to my bank account”). Other user inputs related to medical and health requests also did not work well. Both P1 and P10, who have a tracheostomy and often need suction to clear their breathing pathway, tried using “suction” to generate a request but the model suggested something unrelated like, “I have a suction cup that needs to go on the wall”. Many help requests that users wanted to generate were high stakes and needed to be specific. So when users tried inputs as “meds” or “itch”, the model suggested very generic help requests that participants tended to evaluate as less useful and less appropriate.

Additional Conversational Tasks

Participants also shared different ideas about how they would use speech macros. Several participants said they would program them to make specific help requests to their voice assistants (P1, P6). Others would like to have speech macros help with routine tasks like asking for help with self-care tasks (P7) and ordering at a restaurant (P4, P12) or at a coffee shop (P10), as the model could draw information about what type of food the restaurant sells or what type of coffee order the user always asks for. Some participants also shared that they would not mind using the “Expand Reply Macro” to get suggestions on how to answer common questions like “How are you doing?” which can be frustrating to answer again and again (P1). Speech macros that could draw from background information could be used to plan conversations with doctors (P11) if the medical data given to the system was guaranteed to be secured and private (P10, P11).

7.4.2 Learning Input Mappings

While the macros were able to suggest reasonable phrases that matched the questions asked, participants expressed uncertainty on how the system used the input it was given. In other instances, AAC users proposed use of short-hand as input in order to have more control over the generated output.

Participants suggested using specific notation like adding symbols to their inputs to overcome not knowing what implicit tone the model would decide to go with when suggesting responses. P10 suggested using symbols that could hint the model towards having a more positive phrasing: *“typing ‘word +’ would lead to positive responses.”* Overall, participants expressed a preference for reliable and short inputs. P4 explained he had a macro programmed in his device that expanded

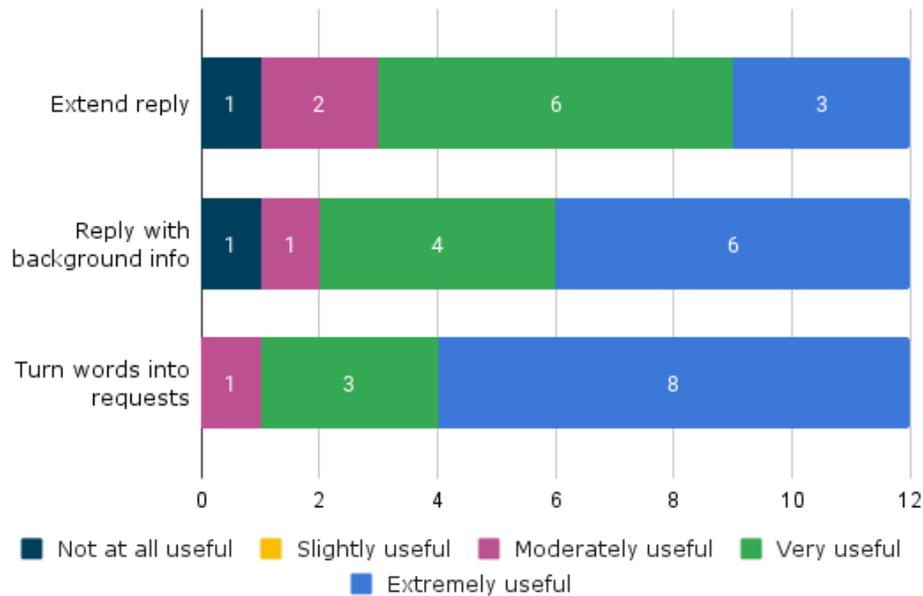


Figure 7.3: Most participants found speech macros to be either very or extremely useful. Turning words into requests was considered very and extremely useful by most participants. Participants identified they could use the turn words into requests macro to help them daily to create requests for routine, self-care and accessibility related tasks.

“1y” to yes and automatically played it out loud saving him time. P8 used “WON1” to say wonderful when others asked how she was doing. P2 used “SYS” as a shorthand that expanded into “see you when I see you”.

An unexpected LLM behavior resulted from having phrase suggestions insert non-factual information, a certain tone, or preference that was not specified by the user. The “Extend Reply” macro would sometimes suggest arbitrary responses that seemed too specific like “I worked from 8am to 6pm” as a response to the question “what did you do today?”. P11 was interested in knowing where the suggestion came from: *“it goes into assumptions of how long you worked. Is that something it learns? That you have patterns?”* Participants were also curious about how to control the variability of the options as sometime the macro repeated phrase suggestions at the testing variability value we tried (0.6).

P10 also noted that some of the phrases suggestions carried an implicit “negative tone” that he would not necessarily intend if he typed the input “work” as a reply to the situation “Do you want to get pizza?” In another case, P3 tried to retrieve neutral answers to the question “how is it going?” by using the phrase “it’s going” as input. The Extend macro did not catch that the user wanted to express a neutral response and instead suggested phrases that complemented the users input: *“it’s going pretty good”*. P3 understood that the underlying speech macro task did not catch her meaning and added: *“if things weren’t going well, I would learn to not confuse the program. and use a different input. [I would use] “not great” for instance.”*

7.4.3 Evaluating Suggested Phrases

While our speech macros performed in an expected way and suggested phrases that were mostly relevant within the conversation scenarios tested, LLM output was insufficient in supporting AAC users in adding their personal tone and style and representing their personality and identity. Participants also shared concerns about how the LLM-generated output could affect their social relationships by being too abrupt or just by the fact that others could know they were automated

responses.

Many phrases were either considered as very appropriate or were phrases that were close to something the participant would say but that needed a little more editing to get it right. Participants suggested ways in which they would use the model's generated output as a starting point to build on top of it during a real-time conversation. *"the first phrases will be like i almost want to answer this it gives me an idea of what i want to say, sometimes but it doesn't necessarily fit exactly with what i want to say... i might end up using a predicted phrase and then delete the last word and write my own"*(P11).

Tone and Style

When discussing the different outputs generated through the three speech macros, participants' most common critique was that the phrases did not reflect their personal style and the image of themselves they wanted to portray. Nonetheless, the way each participant described what qualities of their background mattered to them and what they wished the model knew about them varied. Some participants talked about their personality and how they aim to convey it through their words. Other participants shared the importance of their culture, including their country of origin and their faith and how that impacts the way they talk. Lastly, participants talked about the impressions they wanted to leave on others and how they were concerned the generated phrases could affect their relationships.

Communicating personality, style, and identity

While AAC users found value in being able to get phrase suggestions by typing less, they indicated they would not necessarily use these phrases as they lacked their personal style and did not reflect personality. In many cases, users found phrase suggestions to be "too bland" and impersonal while their personality was "witty" (P2) or "positive" (P8). When sentences lacked a way to convey an important part of the participant's personality and their values, participants said they would not use them. P8 and P10 talked about their faith and personal belief and how that influenced the way they talked to others. *"For me I would add something positive...because I think that positivity in the world is lacking these days so I add positivity into my answers"* (P8). Participants also expressed that they did not wish to sound too scripted. P9 was concerned about sounding "robotic" and P12 raised an important future risk: *"If the system is being used in the future, are all AAC users going to talk the same way? That's something we need to think about."*

Maintaining social standing

Participants expressed concerns related to how the generated phrases could impact their personal relationships. Some participants commonly found generated phrases to be too abrupt (P6, P8, P9, P10, P12) when suggesting possible replies to social questions like "do you want to get pizza?" in the Extend Reply macro scenario, and said that the phrase suggestions were not appropriate for this reason. Participants more commonly preferred phrases that were more socially correct like, "I am sorry, I can't go out this afternoon" rather than the more abrupt ones: "I'm already at work" (Table 7.2). P12 explained why abrupt responses, even though they matched the question, were not appropriate: *"it's not very appropriate because if someone asked me that question and I typed in work, I want phrases like, what time do you want to go? Or I'm working, could we reschedule or give me a minute and I will get back to you about lunch?"*

Participants also had concerns about the system suggesting the wrong thing and making the participants look bad. P4 worried the system would suggest phrases that would reveal information about how he truly felt about someone, or if he had talked badly about someone with somebody and the potential for the system to reveal that fact, embarrassing him. P2 and his wife also had concerns about inputting sensitive words to the system and then having it show unwanted output.

Table 7.3: Participants suggested different items to be turned into help requests. Items spanned three main categories: medical, accessibility related, and daily tasks.

	User input	Participants
Daily living	TV	P2
	Outside	P2
	Sleep	P2
	Bed	P2
	Bathroom	P2, P6, P12
	Window	P3
	Service	P4
	Cigarette / Smoke / I want to smoke	P5
	Tea and Biscuits	P5
	Drink	P6
	Book / Book pages	P8
	Starbucks	P10
	Water	P12
Accessibility	Access	P7
	Transfer	P3
Medical	Suction	P1, P10
	Itch	P1
	Meds	P11
	Pain	P3

P2 wanted to use “bed” to have the model make requests related to helping him prepare for bed. P2’s caregiver was concerned about what the system would say and so did not want him to use that word so they decided to try the word “sleep” as input instead: *“He usually says I’d like to get ready for bed and I didn’t want to use bed because I didn’t want to see what would come up...”*

P2 and his caregiver also shared that for more personal requests like going to the bathroom, they might not want to use the word “bathroom” as a user input but instead a euphemism. *“He would probably use other words at home, and say other words in public to keep it private.”* P2 shared he would say: *“I have to powder my nose.”*

7.4.4 Concerns About Using AI Suggestions

Privacy and data concerns

“ I would love a local version of this [reply with background information macro] which i don’t worry about privacy that i could put medical history information into and it could help me make sure i actually give doctors details when i am in a high stress situation and everything is going so fast i can’t keep up” - P11

Sharing background information with a system was considered extremely useful by most participants. Participants said they wanted to use this feature but would be concerned about other people using their data without their consent, identity theft and data breaches, having other humans reading their personal or medical background information, and on-cloud data processing. Participants asked whether there was a way for them to confirm their data was encrypted and how they could turn off the system from hearing the conversations all the time, as it would need to do this to catch the current dialogue. Participants understood their data and conversation data would improve the model and that the speech macros would indeed be very useful but if there were no privacy in place or clear transparency about how it would work, they would absolutely not use it.

AI suggestions could undermine autonomy

Selecting a generated phrase from the system could have social consequences beyond what the content of that phrase is and how it is interpreted. Participants reflected on how selecting an automated phrase, even a pre-stored phrase they had created beforehand, made others believe the system did all the work for them. P8 shared a story where using her AAC device for a job interview allowed her to get the job because people understood that she had prepared answers to the interview questions beforehand using her device and attributed her with being responsible and well prepared. Unfortunately in another occasion, people thought the opposite: *“I pressed a button for a customized answer and someone said ‘oh, the device did that for you’. That was insulting to get that answer. They thought I was an idiot because I had these customized phrases pre-installed. I was saddened. I had to stand up for myself, so I said I had made the preparation for the meeting.”*

P10 shared that he only allows very few intimate people to watch him type as he communicated and he feels that if they see him select an automatically generated phrase instead of typing his own that could have a negative consequence or change the meaning of that intimate communication.

“... AI generated responses may become the answer for the types of communication that are centered around content and timing. But there are more intimate forms of communication where AI might get in the way of personal expression. I have found with my wife, daughter, and a few close friends that sitting beside me and watching me type the reply to their comment (clearing my response without speaking it) is a secure form of communication for them with me. I would go so far as to say, if the exact expression desired popped up on the list, choosing it would mean something different to an observing intimate friend than if I were to type it.” - P10

Similarly, P8 shared she would not use speech macros with people close to her: “[*I would not use this*] with my family and friends who know me personally. I think sometimes the human feelings cannot be translated by devices.”

7.5 Discussion

In this work, we presented AAC users with an interactive prototype in which they could evaluate text suggestions produced by a large language model. Based on this work, we reflect on how language models may be integrated into AAC devices and about our experiences integrating AAC users into the research process.

7.5.1 Can Speech Macros Improve AAC Use?

Overall, our participants were excited about the possibility of using AI to improve their AAC systems. Participants were clear about the amount of effort they expended when communicating, and the value in reducing some of that effort. At the same time, participants already had ideas about how to improve output from the system, to more clearly reflect their own preferences and communication styles. They also presented some concerns about sharing data, and about the potential loss of *control*.

One question that we considered throughout the work is whether it is useful to focus on specific use cases, as we have done here, as opposed to a more general system that predicts phrases across all conversation contexts, as in some related projects [27, 138]. As a research tool, we found that this approach was successful in introducing participants to the relevant concepts, and providing specific contexts in which to test and evaluate LLM output. Participants also gravitated toward specific macros (especially “Turn Words into Requests”), which may help to prioritize future work. However, whether speech macros should be introduced into AAC devices remains an open question. Existing AAC devices often have multiple modes, such as a mode to replay stored phrases, and it is possible that specific speech macros could similarly be incorporated into current AAC user interfaces. Alternatively, it may make sense to provide AAC users with the ability to create and customize their own prompts to the LLM, enabling them to customize output through prompt programming.

7.5.2 Design Challenges and Trade-offs

While our prototype provided an intentionally simplified interface, creating and testing the prototype revealed tensions between the potential benefits of integrating AI and potential negative effects which we summarize here.

Reducing effort vs. maintaining control

As noted by the majority of our participants, communicating using their current AAC devices can sometimes be both frustrating and ineffective. AI generated suggestions offer the possibility to reduce the amount of effort. However, participants were often unsatisfied with the output from the system, finding it had the wrong tone or was simply incorrect. Participants noted that in some cases they could edit the response to get the result they wanted, but in others they would need to rewrite the entire response. In either case, editing or rewriting an input is counterproductive to the goal of reducing keystrokes. Nonetheless, while our participants wanted to type less to save physical and cognitive effort when responding to routine questions, participants were concerned about how automatic phrase generation could impact their relationships. Participants shared how putting effort into their communication by preparing long messages demonstrated to others that they cared, about a job interview (P8) or about a close family member (P10). These findings align

with prior work highlighting how effort invested in computer-mediated communication can be a symbol for caring [79]. While keystroke savings can reduce time and effort, future explorations with AI should consider how views on authorship and effort may impact relationships among AAC and non-AAC users.

When encountering low quality suggestions, participants sometimes tried to enter longer queries, or invented new input conventions, such as entering a plus after their input to retrieve positive responses. Providing more robust input options could provide users with more control. Similarly, showing users potential predictions as they type could help users make decisions about when to try predictions vs. typing out messages themselves.

Composing in real time vs. using stored content

AAC users optimize and plan for their communication and social interactions. A lot of preparation happens, they create and store phrases that anticipate potential questions or any high pressure situations they may encounter. While this study does not engage with situations in which AAC users could use the technology to plan content for future encounters, participants gave clear feedback of how the output could be better and even how they wanted to combine phrases created by them with the ones generated by the system.

It is important not to ignore all the work and setup AAC users have already established that work for them, like the shortcuts they already have in place (i.e, 1y for yes, WON1 for wonderful, etc). If AI will be used within an AAC device, it should be flexible so that it does not impose a scheme but learns from and is customized by the user who has already developed a system for their communication.

Achieving functional communication goals vs. expressing oneself

Prior work has reported that the way people communicate changes once they start using AAC systems [77, 102]. Individuals who acquire speech disabilities later in life often lose the ability to express sarcasm, humor, and nuance. The social timing pressures that exist for AAC device users make it hard to add nuance and talk about other things beyond requests or short utterances [78]. This is why sometimes there are large efforts into enabling basic communication with AAC devices, and even though there has started to be a shift into how AAC devices could also play a role to maintain social relationships [35], we did not know what qualities of generative output could be important to know about for future systems. Our study revealed that AAC users want more from LLMs in addition to keystroke savings and achieving that model-generated golden reply that is reasonable within context. Users want to be able to customize output to their needs (and this is different for each person); this is the key to unlocking the potential of LLMs. Moving away from scriptedness and transactional conversation support towards customized use of these systems.

While prior work established that computer-mediated communication (CMC) grants greater control over the impressions people convey to others as people can edit and plan their messages before sending them [162, 59], AAC communication is a unique type of CMC where responses are expected sooner, and the time window to achieve self-representation is shorter. Perhaps by enabling post-processing of LLM suggestions or co-authoring of a response we might support AAC users in personalizing their responses on the go.

Additionally, our study revealed the importance of customized user information to better tailor LLM-generated output to. Prior work created simulated personas mostly comprised a one to two sentence description of a person's hobbies or personal preferences to generate conversational phrases [138]. Through our study we gathered other important information about what AAC users would like a model to know about them: medical details, details about their relationship with conversation partners (co-worker, family), details about their work and context information about their location or time of day.

7.5.3 Conducting Human-Centered AI Research with AAC Users

Conducting participatory and open-ended research with AAC users can be challenging as new methods are always needed to elicit real-time feedback in a way that aims to maximize participation and reduce user burden [8, 17]. These challenges are amplified by the need to conduct studies remotely during the COVID-19 pandemic. Despite these challenges, participation from AAC users is necessary to ensure that technology accurately meets their needs. A primary goal of this research was to enable AAC users to interact with, and provide feedback about, a real AI model. All AAC users were able to successfully suggest inputs and provide feedback about generated outputs. Additionally, the use of a post-study survey was successful in allowing participants to share more extended thoughts, anecdotes, and feedback.

We conducted each of our 12 interviews remotely, and encountered some anticipated and unanticipated challenges while doing so. First, we knew that the pace of interaction during the study might be slow, and that the technologies we wished to discuss might be unfamiliar to our users. Second, we expected that some participants would be limited in their ability to give feedback during the session. Third, because most participants would be using their devices for communication, we could not install software on their devices.

Generally, we were able to adapt our study protocol to these circumstances. First, we designed our three speech macros to be simple and easy to understand. We designed our prototype to quickly show the language model by allowing the user to change the contextual information or the prompt and quickly see the results. Through building in a variability slider we were able to run the model at other randomness values if needed. By using screen sharing, we were able to accept spoken or typed user input through chat and input it to the prototype. Finally, as in previous studies [77], we combined our in person interview with a follow-up questionnaire so that participants could compose longer responses. Our study allowed us to explore potential uses of AI for AAC by enabling users to directly interact with a language model via digital prototypes. While our speech macro concepts served us to communicate LLM’s capabilities and usage scenarios, co-designing future speech macros or prompts with AAC users may enable us to understand how AAC users create mental models and expectations for LLMs.

Overall, the design of our prototype and study protocol enabled our participants to see the LLM in action, test it with several queries, and provide feedback during and after the study session. While this configuration worked well for this study, conducting longer deployments of this technology would require an alternative setup.

7.5.4 Limitations and Future Work

While we designed our prototype and study to maximize the amount of interaction participants would have with the language model, they were still limited to completing 5-10 inputs in the study session. As a result, participants gave feedback based on this limited experience only. Future studies could feature a longer deployment, so that participants could input more prompts and gain a better sense of the strengths and weaknesses of this approach.

The language model used in the study was designed for generic dialog-based applications; each speech macro was a prompt written for that specified activity. We did not customize speech macros with any data about specific users. While the speech macros proved effective for collecting user feedback, this study does not provide a clear picture of how well current language models could perform for AAC suggestions, and it is likely that tuning the speech macro prompts or adding user-specific data to the prompts would improve suggestion quality. Furthermore, participants provided suggestions for new macros; providing end users with the ability to edit and customize their own macros could further improve results.

We designed our *speech macros* prototype with the goal of quickly introducing users to the concept and enabling them to try out several macros during the study session. By design, the prototype does not include details of user interaction, such as whether suggestions should be added

to an on-screen keyboard or placed in a menu, or what actions are needed to generate suggestions. During our study, participants emphasized the importance of having control over predictions and the usefulness of composing messages ahead of time. Participants were also aware that the system might not produce the output they intended, and that they might need to fall back to typing a message themselves. Future work could explore the design of appropriate user interfaces that combine contextual information, saved settings, and live input, and allow users to correct or override the system if it fails to produce usable output in a particular situation. There may also be ways to provide users with more control without requiring more typing, such as by allowing users to specifically request a positive or negative response, or to request short or long responses.

For this study, we recruited participants who used AAC due to motor difficulties; our participants had typical language skills and did not have conditions such as aphasia that would impair their understanding of spoken or written messages. We were thus able to assume that participants were able to understand the purpose of the macro and its expected input, correctly format their input messages, and choose from the suggested phrases without difficulty. We expect that this approach could be made useful to those with aphasia and other language disorders by designing appropriate interfaces to detect and highlight potential issues related to comprehension, or by personalizing user models that accept input and generate output in a format more appropriate for that individual user.

7.6 Summary

We conducted an evaluation of large language models as a tool to support AAC users in generating phrases, introducing *speech macros* as a method for AAC users to benefit from the generative capabilities of these models. Our study found that AAC users were enthusiastic about the potential of language models to support their communication, provided that they maintain control of their personal expression. The potential of LLMs for different types of AAC use should be explored through future design work and experiments.

Chapter 8

Augmenting Context Grounding with the COMPA system

8.1 Overview

Grounding in conversation is the process of establishing mutual understanding. One way interlocutors achieve grounding is by ensuring they know what each person is talking about or referring to. To facilitate grounding in conversation, interlocutors tend to add information to their own utterances by first establishing the identity of what they are referring to, then further elaborating (ex. *“The dog, he just bit me”*) [30]. Nonetheless, using references (i.e., “talking about X, I think...”) is a costly and difficult grounding strategy to achieve in AAC because the mediums of communication are asymmetrical: one person is speaking at fast rates while an augmented communicator is typing their message at considerable lower speeds [136, 30, 158].

Albert Robillard, a researcher who studied action and talk in situated behavior and used AAC methods to communicate explained not being able to repair his utterances in the intensive care unit [125]. He explained that he was often cut-off mid-sentence when someone left the room before he had finished his message, as he could only communicate a few words per minute, and when he wanted to resume speaking he experienced what he called the “Out of Context” problem: *“Usually the interaction has moved along so far that when I address an old topic my conversants have a hard time seeing the relevance of what I am saying. It takes so much effort to spell out what I am saying I could not easily recycle the topic by saying “You know what we were speaking about a little while ago, the X topic.” I could only, because of time and energy, speak directly to a former topic. The speaking out of context would generate many complaints and confusion.”*

Therefore, it becomes even more important to find intuitive solutions to address the out-of-context problem for communication between AAC users and non-AAC users.

Grounding is dependent on the medium of communication [30] and little is known about how specific AAC interface that designs may impact establishing a common ground. In this chapter, we propose different ways to support establishing mutual understanding and common ground despite time differences in the response time among AAC and non-AAC users. In this chapter, we explore how different conversation grounding strategies might loosen the relevance constraint for AAC users and their communication partners. We introduce COMPA, a system that incorporates different types of contextual grounding strategies such as highlighting and alerting a communication partner about the current conversation portion an AAC user might be composing a message about. COMPA also integrates a large language models that provides starter phrases as suggestions that can help re-introduce a past topic into conversation. This Chapter describes the design process of COMPA, its features and an multiple condition study in which we sought to understand how different COMPA components contribute to supporting AAC users in having a better mutual understanding.

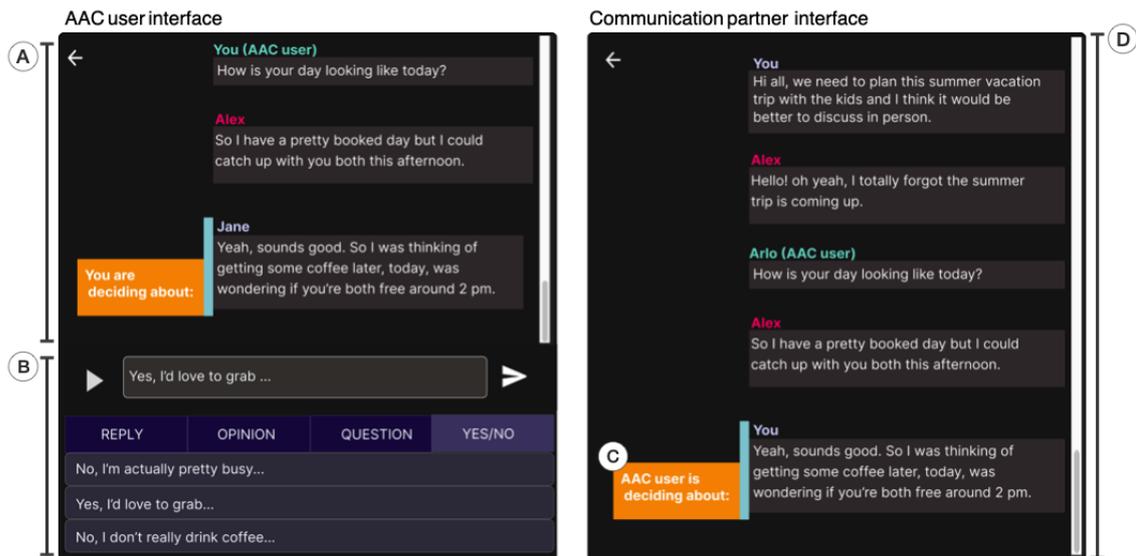


Figure 8.1: COMPA is a browser extension that enables Context Marking and Phrase Assistance to augment contextual grounding and awareness about what an AAC user may want to refer to in conversation. Using COMPA, AAC users and communication partners can view a live transcript (A,D) of their conversation and notifications (C) about the context of a conversation an AAC user may want to reply to. COMPA also explores how enabling AAC users to choose an intention when composing a message (*i.e.*, share an opinion or ask a question) can provide better starter phrase suggestions and more accurate communication partner notifications (B).

8.2 Design Goals

When AAC users are typing co-occurring talk by other conversation participants may change the topic of the conversation faster than what AAC can respond to [44, 66, 158]. Sometimes this means that their messages might be taken out-of-context, or need additional clarification from an experienced partner to be re-contextualized. In response to the out-of-context challenge and in order to loosen the relevance and timing constraints, our studies in chapter 6 and chapter 3 indicated possible design directions to support three goals (G1-G3):

G1: Share Conversational Context. In the conversations we analyzed in chapter 6 and 3, communication partners (CPs) often clarified messages for the AAC user when these messages were delayed and ambiguous responses. Often CPs guessed what the AAC user meant to say but other times they needed to confirm with the AAC user that their clarification was correct. As such, supporting having a shared conversational context such is an important design goal to help AAC and non-AAC users have a clear understanding of the part of the conversation a person may be referring to. Additionally, during the ideation workshop (chapter 6), an AAC user shared how AAC devices are not currently responsive to what is happening in a real-time conversation and how having the device adapt to the ongoing conversation could better support him in communicating in a timely manner. We then consider the design goal of using the conversation as a source of information that can provide relevant phrase suggestions tied to the ongoing conversation.

G2: Provide Awareness of AAC User Intent. As explored in the workshop, knowing the intention or stance a person has towards a topic can help others better understand their contributions in conversation. Having a way for the AAC user to specify their stance or intention towards a topic is an important design goal that could help their CPs provide better support and explanations on their behalf when needed, or even improve the word and phrase suggestions the device provides. Making an AAC users intention salient could also create more specific cues for their conversation

partners.

G3: Support AAC User & Partner Initiated Alignment. Communication partners who know the AAC user well often act as mediators between unfamiliar partners and AAC users. Evidence of this is thoroughly describe throughout this dissertation. Similarly, as also mentioned in the interdisciplinary ideation workshop (Chapter 6) AAC users' participation in conversation would greatly improve if others just learned to wait and slow down. Tools designed to support mutual understanding between AAC and non-AAC users should be shared, similar to groupware explored in prior work [46], out third design goal is to create AAC and partner interfaces that can support their mutual adjustment and learning process to communicate with each other.

8.3 COMPA

We introduce COMPA, a system to support creating a shared understanding of conversational context, and awareness of AAC user intent (Figure 8.1). To support shared conversational context, we explore displaying a **real-time transcript** of the conversation. We enable AAC users to **pause** the transcript while typing their response to help partners understand the context of their response (Figure 8.1, chat panel). We **notify** both the AAC user and the partner when the transcript has been paused. We leverage the transcript to create **phrase suggestions** based on the current conversational context to let the AAC user easily *add context to their comment*. To support awareness of AAC user intent, we enable AAC users to also select an **intent** (reply, opinion, question, yes/no). When the AAC user selects an intent, we add the intent to the pause notification to help the partner understand the desired action that the AAC user wants to take in the conversation. We also enable the AAC user to leverage the intent to generate phrase suggestions based not only on the conversational content, but also on the specified intent.

We implemented COMPA as a browser extension built using React.js to support AAC users and their communication partners participating in remote meetings. We selected remote meetings as a first use case for COMPA due to the availability of high quality real-time audio recordings from computer microphones to enable transcription, and built in access to a shared screen to support mixed AAC user and partner focused conversational support.

8.3.1 Interfaces

COMPA has an **AAC interface** and a **CP interface**. Both the AAC interface and the CP display the transcript of the current conversation on the transcript panel. The AAC interface additionally includes a text composition panel that enables the AAC user to compose a message and use text to speech to read the message out loud.

Transcript Panel

The transcript panel displays the conversation speakers and live transcriptions of the ongoing dialogue. The AAC user can pause the transcript panel by starting to type or by pressing a pause button. This pause action allows AAC users to pause the transcript at the point at which they want to speak. After they finish speaking, the tool will insert their utterance at the point in the conversation where they paused, and any conversation that has happened since then will populate below their utterance. The transcript on the CP's tool will pause and continue the same way.

On both the AAC user and CP's interface, a set of previous utterances in the conversation will be highlighted to **mark a context** the AAC user may be speaking to and to notify the CP that the AAC user is typing (Figure 8.1). This serves to further ground what the AAC user is responding to and establish the relevance of their future message. The AAC user and CP see the same context

Table 8.1: Our eight study participants included four AAC user study participants (2 females) and four corresponding communication partners. AAC users have 10 or more years of experience using AAC and use a diverse set of AAC devices and access methods.

AAC User	Age	AAC Device and Access Method	Experience (years)	Communication Partner
AC1	38	Tobii I 15 with single switch	32	CP1, Sister
AC2	53	Various text-based AAC; direct access by typing with fingers	20	CP2, Daughter
AC3	58	Speech Assistant app via an EMG switch; NextUp Talker via a SCATIR switch and a Words+ EZ Keys dongle interface.	10	CP3, Spouse
AC4	45	iPad with Predictable and Speechify apps	40	CP4, Sister

mark in their own interface views. The tool currently highlights the last or current turn closest to the time that the AAC user presses the pause button or starts typing.

Text Composition Panel

The text composition panel consists of a pause button to control the transcript panel view, a text input message window where the AAC user can type using their AAC device or preferred keyboard, and a submit button that will read the AAC users message out-loud and print it on the transcript panel after it is activated.

COMPA provides **starter phrase suggestions**. Starter phrases are defined as a sentence fragment that starts off a valid response based on the previous context of the conversation. These phrases are generated by prompting openAI’s ChatGPT (model gpt-3.5-turbo) [110] with manually created examples of starter phrases to various conversations, then providing the context of the current conversation for ChatGPT to start responding to. Three phrases are generated at each time the AAC user starts typing below the text input window (Figure 8.2 and Figure 8.1). We provide three phrase suggestions as to not overwhelm the AAC user with many options while still providing a variety of possible responses.

COMPA also includes **intention** buttons an AAC user can select to indicate how they want to respond in the conversation. This intention buttons are informed by common communicative intentions or dialogue acts commonly used in conversation [115]. We condensed the intention into four groups: reply, opinion, question, and yes/no. This is done to not overwhelm the AAC user with more than four choices. These groups are chosen as most responses can fall into one of these categories: reply (ex. statement-non-opinion), opinion (ex. statement-opinion), yes/no (ex. yes answers, agree/accept, response acknowledgement), and questions (ex. yes-no-question, open-question) [75]. The AAC user can select an intention prior to their response that will affect how starter phrases are generated. Once an intention is selected, instead of the generic message of “(the AAC user) is typing about”, the CP will see a message such as “(the AAC user) has an opinion about”. An exhaustive list of these messages can be found in Figure 8.2 and Table 8.2.

8.4 User Study

We designed three versions of COMPA (figure 8.2) to study how different features and changes to each interface’s panels and different grounding strategies may impact communication for AAC and non-AAC users. Each version of COMPA provides different context grounding strategies, phrase assistance types to the AAC user and different types of feedback notifications to the CP, (Table 8.2):

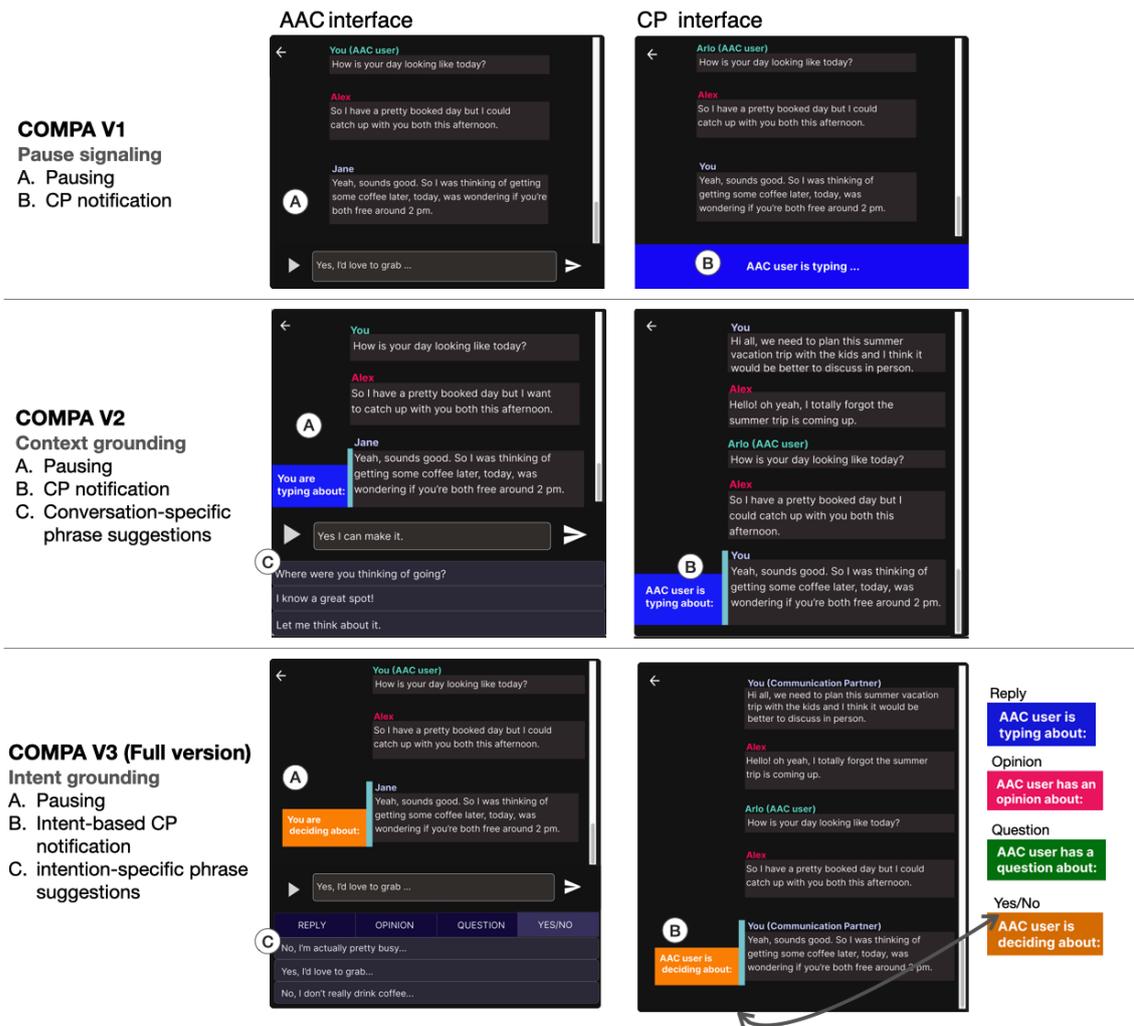


Figure 8.2: We created three versions of COMPA. Each version added a new layer of context information and phrase assistance type. Testing these three different versions would allow us to evaluate the contribution of each grounding strategy to the conversation experience.

- Version 1 (Pause Signaling): contains only the context grounding functionalities, pauses the transcript and displays general typing notifications to the CP.
- Version 2 (Context Marking): contains starter phrases for the AAC user that are conversation-specific in addition to the context grounding functionalities.
- Version 3 (Intent grounding, *full version*): contains context grounding functionalities (pauses and CP notifications) and intents that the AAC user can select and intent-specific CP notifications in addition to the starter phrase that are conversation and intent specific.

By having these three versions of COMPA, I wanted to understand how each feature impacted AAC users' and CPs' participation and communication. I had two initial hypotheses:

- **H1.** Adding more context and intent information will help AAC users communicate more clearly and more often.
- **H2.** Adding more context and and intent information will help Communication Partners have greater awareness of the AAC user's participation.

8.4.1 Participants

A total of eight participants, four AAC users (2 Women) along with a communication partner (CP) of their choice were recruited for this study (Table 7.1). Each AAC user and their CP were given a pre-study survey, which asked about each participant's communication strategies to their partner as well as each participant's perceived participation time during group conversations.

We conducted a two-hour remote study session for which we connected via Google Meet. We asked each participant, AAC user and CP to connect from their own computer and from a separate room if they were in the same household to minimize any text-to-speech error. A researcher and author of this work acted as the study moderator. Using a within-subjects experimental design, each pair of participants used all three versions of COMPA. We counterbalanced the order of the conditions to prevent acclimation bias.

8.4.2 Procedure

The remote user study consisted of four main parts for each interface condition: (1) an overview of COMPA's interface version, (2) a short tutorial, (3) a conversational task: planning a summer vacation, and lastly (4) an online questionnaire post-condition.

Tutorial

After a general explanation of the features, the moderator performed a storytelling tutorial task for the pair of participants so that they are familiar with the version's interface. The storytelling task consisted of the moderator asking the AAC user to use the features of the corresponding COMPA version to respond with the word "green" and a selected phrase starter (when applicable) upon hearing the same "green" keyword in the story narration. The CP is also invited to monitor their COMPA view and describe the changes they perceive on their screen.

Conversational tasks

After the tutorial, the moderator, AAC user, and CP participated in a structured conversational task of planning a summer vacation together. This task was split into three sub-tasks: (1) decide on the location for the vacation, (2) decide on the trip activities, and lastly (3) discuss who else

Table 8.2: Participants tried three COMPA versions that had different levels of context information for each of COMPA’s features. Participants tried all versions in a randomized counter-balanced order.

COMPA Version	Transcript Behavior	Starter Phrase Assistance	CP Notification
1. Pause Grounding	Pause	None	”user is typing”
2. Context Grounding	Pause + marks context	Conversation-specific	“user is typing about:” [context]
3. Intent Grounding	Pauses + marks context + user selects intention	Conversation & Intent- specific	”user has [intention] about:” [context]

to invite to the trip. The group completed each sub-task with a different version of COMPA. The order of these sub-tasks remained constant during each run of the study. The summer vacation task was chosen to simulate a natural conversation among a group where different sub-topics could be discussed under these big theme of vacation planning and these clear tasks of choosing a location, activities, and people to invite.

Each sub-task lasted about 10 minutes and the moderator guided the conversation in a specific structure to maximize uniformity across participants. First, the moderator instructed participants on the items to be discussed (location, activities, or people to invite) and then proceeded with a direct question to the AAC user to hear their input first. Then, the moderator had to ask a direct question to the CP and make at least one general question to the group. The moderator also had to create one big topic change by inserting an unrelated anecdote to the conversation or asking a question about a new topic. The moderator also had to make at least one common error unfamiliar partners make with AAC users such as speak while the AAC user is typing or ask two questions in a row. These moderator behaviors were selected to try and reproduce possible errors an unfamiliar partner might make.

Post-condition surveys

Proceeding the use of each version of the tool, the AAC user and CP received different post-condition surveys. The AAC user was asked how their experience was using the tool, as well as how effective they thought the tool was in helping them communicate or participate more. The CP was asked questions targeting their awareness of the functionalities on the tool and their ability to support their partner. Each were also asked about the specific functionalities in the version of the tool they are using.

To conclude the study, the AAC user and CP received the same set of questions that ask them to rank the three versions of the tool in order of usefulness for them and for their partner. As a take-home survey, they are asked how likely they would use this tool in meetings and in what other scenarios they could use the tool.

8.4.3 Analysis and Data Collection

We recorded the remote study sessions (video and audio), the interaction logs of COMPA’s interfaces, and survey responses from both AAC and CP participants. We reviewed the data to extract both subjective and objective metrics. The subjective metrics included perceived sense of agency, that is if participants felt they participated in the conversations as much desired; perceived ease of input or effort required, and perceived awareness of screen changes of the CP. Objective measures included the number of turns taken by each participant and length of AAC user contributions per

turn (words used per turn). To understand how the different COMPA features impacted the AAC users' participation we also labelled the type of communicative functions used during each AAC user's turn (e.g., giving an explanation, confirming information, requesting help). We compared these metrics across the three COMPA versions. These metrics are all factors that can impact the effectiveness of communication and mutual understanding.

8.4.4 Study Limitations

While the study lasted two hours and included tutorials for each version of COMPA, participants in this study all experienced COMPA for the first time during the study. Thus, it is possible that with long-term use of COMPA our results may vary. Similarly, our conversational task used in this study to elicit a natural conversation, sharing, and collaborative decision-making was limited to one big theme of planning a summer vacation. Future work is needed to understand how different COMPA features may or may not support different types of conversations (i.e., personal, work-related, etc.).

8.5 Findings

We first report on participant's experience using COMPA, the amount of times they participated and the type of phrase suggestions they used. We then report on participant's feedback about each of COMPA's features and lastly we report on the observed participation behavior and subjective participation metrics that describe participants technical agency, or ways they participated in conversation and their perceived sense of agency across versions.

COMPA was successfully installed by all study participants. Nonetheless, AC1 had trouble accessing COMPA directly with their AAC device as it restricted Google Chrome browser access. AC1 and CP1 still completed all study tasks and completed a take-home survey to share their opinions on the COMPA's feature. We only include AC1's and CP1's feedback and ratings but not their participation metrics information.

8.5.1 Using COMPA's Features in Conversation

Participants rated COMPA's features to be in general very useful to them (Figure 8.3). From all of the versions they tried, AAC users selected COMPA's second version more often as the version they would most likely use while CPs selected the full version of COMPA as their preferred version (Figure 8.4). Participants shared they really liked that COMPA worked as an add-on that can be accessed with their AAC device and that included text-to-speech. ACs shared it is often hard to switch back and forth between remote meeting software and their own AAC device, but because COMPA integrated text-to-speech and is connected to the meeting it really helped overcome this obstacle.

Transcript Pausing. AC3 shared that the transcript pausing feature, "*seemed more of a level playing field.*" AC2 appreciated the transcript feature as well: "*As a person who speaks English as a second language, I appreciate the transcription feature embedded in the AAC app.*"

Having access to the transcript of the conversation also highlighted opportunities to include additional functionalities like translation. Two of our AAC participants were bilingual and both suggested embedding translation features into the transcription as the transcript would often misspell proper names that were not of English origin.

CPs found it was easy to understand their Augmented partners better with the transcript but shared the conversation was easier to follow when the transcript "marked the context" of what the AAC user wanted to respond to (Figure 8.3, Marking Context Feature).

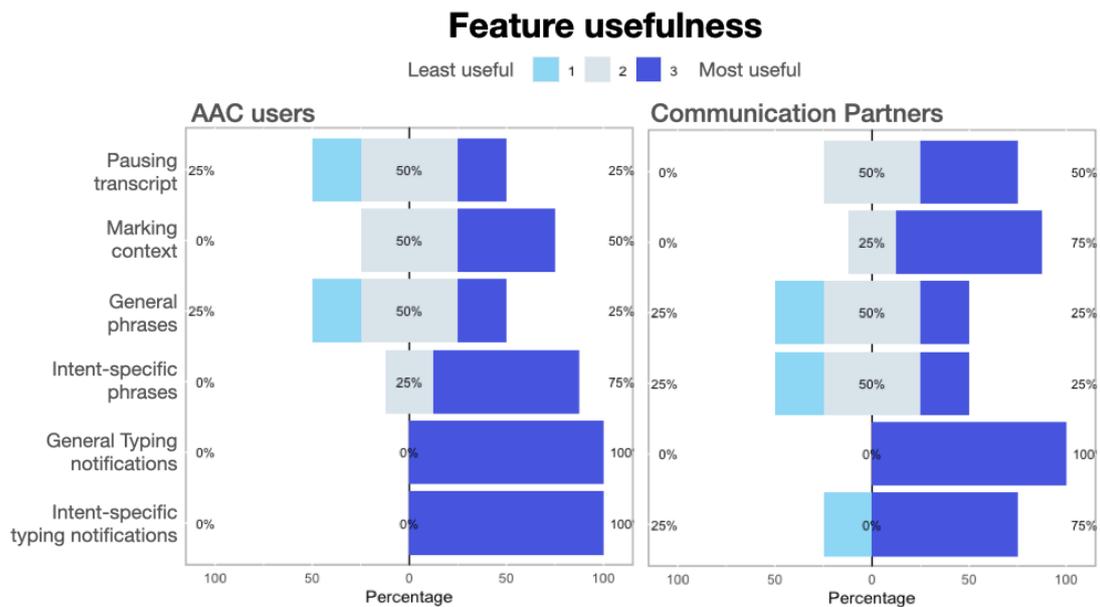


Figure 8.3: All study participants (N=8), four AAC users and four communication partners (CPs), rated all of COMPAs features after experiencing all versions of the tool. CPs rated pausing the transcript, marking the context, and general typing alerts as the most useful features to help them better communicate with their augmented partners. All AAC users rated the general typing alerts and the intent-specific typing alerts as highly useful to them.

Partner Notifications. Communication partners found the intent-specific notifications to help the flow of the conversation. CP3 shared “*the boxes like reply and opinion helped the conversation*”. CP2 also liked having more information about the AAC user’s intention:

“I liked that we could anticipate what type of response she was going to make ”
–CP2

CP4 shared that because she is used to communicating with their sister the intent-specific notifications did not really change their ability to communicate with each other: “*we interacted about the same as usual...I feel the options on how to respond would benefit people who are not used to communicating with someone with AAC*” – CP4

In general all AAC users found partner notifications to be one of the most useful features to them and CP preferred found general typing notifications to intent-specific ones (Figure 8.3).

Phrase Suggestions. AAC users used phrase suggestions to reply in conversation. AAC users asserted their agency by always editing starter phrase suggestions with small modifications. Participants preferred using intent-specific phrase suggestions to conversation-specific ones. In total, conversation-specific phrase suggestions were only used twice. AC4 used the conversation-specific phrase suggestions twice. AC4 used the phrases to reply to two specific questions and modified the phrases using the message input box:

MODERATOR: I was wondering if you all do any research about like, Before you decide to go somewhere. If it’s accessible or if it has some accessible spots before you decide to go?

AC4: *[pauses transcript by starting to type: “y”, then stops and sees phrase they want and selects it. Phrase selected is “I actually like looking for accessible spots.” Then goes back and finishes editing her message.]*

AC4: yesI actually like looking for accessible spots.

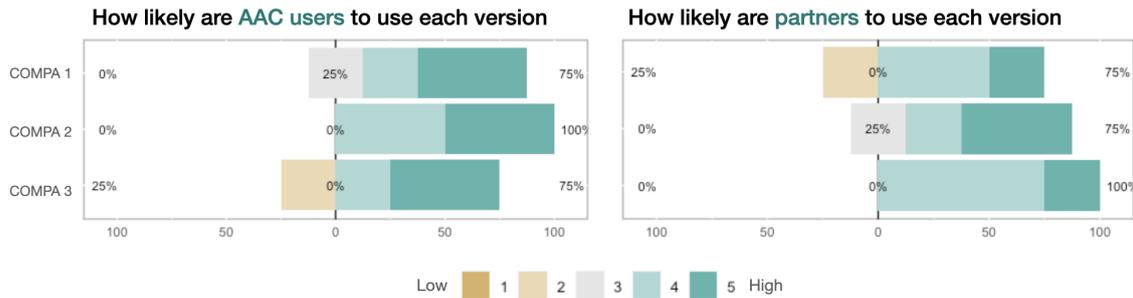


Figure 8.4: AAC users rated each in version in terms of how likely they would use it again if it were available to them. Most AAC users rated COMPA’s second version as the version they would most likely use while CPs selected the full version of COMPA as their preferred version.

Similar to the example above, AC4 used a suggested phrase “that sounds great to me!” to quickly reply in the conversation and modified it by adding “ye” to the beginning: yeThat sounds great to me!

AC2 and AC3 did not use any of the conversation-specific phrase suggestions. AC3 shared: “*I didn’t feel a good fit with the suggestions.*” AC2 shared she did not use them because she “*mostly typed without looking at them.*” AC2 and AC3 preferred the Intent-specific suggestions. Intent-specific phrase starters were highly rated as being appropriate and useful. In total, AAC users (N=3) used five intent-specific phrases (COMPA version 3). Two users shared they preferred the intent-specific phrases, including AC2 who used the intent-specific phrases in 33% of her turns when having a conversation using COMPA 3. AC3 also preferred the intent-specific phrases, “*I actually think communication was more relaxed because of the structure provided.*”

AC2 who used the most intent-specific phrases, selected a phrase and edited it in 3 out of 4 occasions. In one example when the moderator was talking about seafood, AC2 selected the reply intent-specific phrase: *I’m not a big seafood fan, but..* and edited it to “*I m a big seafood fan*”. Other edits included selecting “*I love trying new things...*” and changing it to “*I love trying new food.* ” and lastly selecting “*What kind of food do you. . .*” to “*what specillized foold does your country have?*” (this spelling error is as it was written).

Intention Buttons. AAC users selected the intention buttons 7 times in total to change their partner notification message (see Figure 8.2, for a list of the intent-specific CP notifications). Even though AC4 did not use any intent-specific phrase suggestion, they used the intent buttons to indicate their intention to their partner. The “reply” intention was more often selected than other intentions (5 times in total). The opinion intention and the question intention were only chosen one time each, over all.

8.5.2 Effects of COMPA’s Features on Participation

Amount of participation. In average AAC users took more turns when using COMPA’s full version (COMPA 3) but used more words per turn when using COMPA 2. In general, CPs took more turns in conversation than AAC users. The distribution of turns taken by AAC users and CPs was comparable when participants used the full version of COMPA (Table 8.3).

Quality of Participation. In Chapter 3 we quantified AAC users and CPs conversational turns and labelled the communicative functions they used when participating in conversation to measure occurrences of AAC users technical agency. Technical agency is observed as individual contributions to a conversation. To measure how AAC users were able to express their technical agency in this study we counted the number of turns taken by each participant, the words said per turn

Table 8.3: Average turns taken and words per minute spoken by turn. The average number of turns taken by AAC users and CPs was comparable when participants used the full version of COMPA.

Version	AAC users				Communication Partners			
	AC turns	Stdev	Words per turn	Stdev	CP turns	Stdev	Words per turn	Stdev
1	9.67	2.31	5.00	1.43	7.67	1.528	9.45	1.25
2	11.00	3.00	5.87	0.60	13.33	5.033	10.13	1.78
3	12.00	2.00	5.04	1.50	13.33	2.082	11.15	2.89

and the type of contribution each participant was able to make while using each COMPA version. Contribution types were informed by the metrics developed in Chapter 3.

I developed a coding scheme to label each turn taken by the AAC user and their CP after examining all the video recordings and the session transcripts. I labelled each turn with a “move” and a “function”. Moves are defined as a unit of talk within a turn that has a function [31]. I observed two main types of moves (1) *Replies*, which included any response to a question or another person’s comment and (2) *Initiations*, which are contributions that open the conversation or introduce a topic and usually require a response from others. I further labelled the replies with additional labels: (1) *giving a detailed reply* and (2) *giving brief ambiguous reply*. Additionally I labelled the number of ideas or preferences expressed by each participant (Figure 8.6 as a way to understand how often they were able to share their own wishes during the summer vacation conversational tasks.

Observed Technical Agency. AAC users produced more initiations when using the full version of COMPA than when using version 1 or 2 of the tool (Figure 8.6). AAC users also initiated conversation more often than CPs when using the full version of COMPA. All AAC users were able to contribute at least one idea or preference for the summer trip planning activities.

Observed Context use. Version 2 and Version 3 enabled participants to use phrase suggestions that could add more details to their replies. The findings show that AAC users produced more detailed replies when using Version 2 and 3 compared to Version 1. Even though AAC users contributed more ambiguous and brief replies than detailed replies, there was only one instance, across all participants and all study sessions, where a CP had to clarify a message on behalf of the AAC user. In this case, CP4 asked AC4 clarifying questions to fully understand what they intended to say. For all other participant pairs, communication partners and the moderator were able to easily understand what AAC users meant even when some of their responses were brief and lacked references to a specific topic.

Perceived Sense of Agency. AAC users rated their perceived sense of agency in terms of how often they felt they were able to say what they wanted to say during each conversation with COMPA and how much they felt they participated compared to their usual participation in group conversations. AAC users felt they participated as desired more often when using COMPA Version 2 and felt they participated much more when using the full version of COMPA (Figure 8.5). AAC users perceived amount of participation did match with the observed number of turns quantified (Table 8.3).

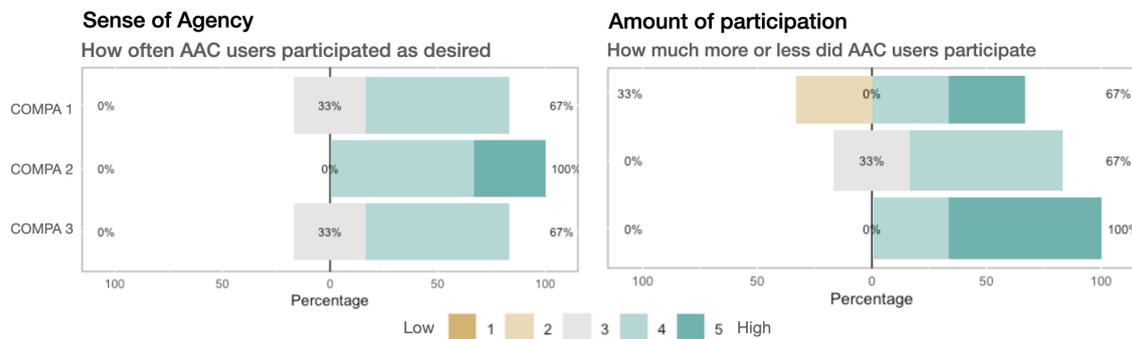


Figure 8.5: AAC users stated they were able to always (5, rating value) or often (4, rating value) participate as desired when using the second version of COMPA. AAC users stated they were either often (4, rating value) or sometimes (3, rating value) able to participate as desired with COMPA 1 and COMPA’s full version. AAC users felt they participated much more when using the full version of COMPA (COMPA 3) compared to the other two versions.

Frequency of Moves and Communicative Functions

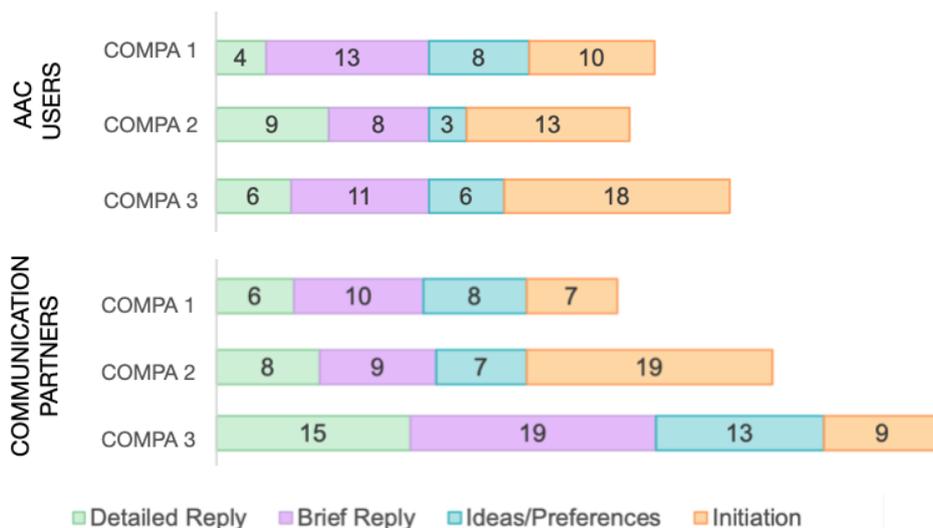


Figure 8.6: AAC users initiated conversation more often when using the full version of COMPA (18 total initiations by AAC users). COMPA’s full version enabled AAC users to retrieve intent-specific phrase suggestions and share intent-specific typing notifications with their partners.

8.6 Discussion and Future Work

This study aimed to understand how different contextual grounding strategies could support AAC users and their conversation partners in conversation. We created COMPA with specific features that could support three design goals: (1) Supporting sharing a conversational context, (2) Provide awareness of AAC user's intent and (3) Support AAC users and partners in achieving mutual understanding and alignment. COMPA's design goals are agency-driven goals, informed by the prior chapters and derived from our Agency Framework.

Partners increase agency for AAC users when they know more about the conversational context, so COMPA aimed to provide both AAC users and Partners with more information about the current conversation in form a transcript that pauses when the AAC user types and partner notifications that mark a context an AAC user wants to speak to.

Additionally, we had learned in prior chapters of this dissertation that communication partners can benefit from more signaling to be reminded of when and how AAC users want to participate in conversation. COMPA aimed to provide awareness of when the user wants to participate and how they wish to participate by adding the intent selection feature. COMPA also provides starter phrase suggestions via a language model that consider the ongoing dialog between the meeting participants and the intention selected by the AAC users. Chapter 7, revealed that complete AI phrases may not adequately reflect what an aac user wants to say. COMPA suggests "starter" phrases instead that can be edited on the go to quickly to re-use context embedded in them but also add the users' personal touch.

Lastly, we know that agency is socially distributed, AAC users and partners cooperate with each other in conversation to achieve mutual understanding. So we wanted to design COMPA to support both AAC users and partners. Communication partners were asked to monitor their screens to track notifications and the conversation's context represented by the transcript.

All of these agency factors resulted in COMPA's three grounding strategy ideas, conveyed through COMPA's different versions (Figure 8.2): (1) pausing a shared live transcript of an ongoing when the AAC user starts typing; (2) have conversation and intent-specific starter phrases generated by a LLM that can be used to recontextualize any comment an AAC user wants to make in conversation; and (3) communication partner notifications that can be either general or intent-specific.

Different COMPA versions were evaluated with three conversational tasks about planning a summer vacation between an AAC user, a close conversation partner, and a moderator (myself). The moderator's role was to guide the conversation but also introduce topic changes and possible noise by talking while the AAC user typed, participants were able to communicate effectively and did not have any major misunderstandings across all tasks and conditions of the study. All communication partners found it useful to know when AAC user where typing and 75% of them found it was useful to see what specific part of the conversation the AAC user was typing about (context marking feature).

Although AAC users took less turns than communication partners (CPs) when using all COMPA versions, the number of turns AAC users took was in average very close to the number of turns taken by CPs which is a possible indication of a balanced interactions where all conversation participants participated almost equally. When examining the communicative functions and moves used by AAC users across the study conditions, we see that AAC users initiated conversation much more when they had access to the full version of COMPA. Starter phrases enabled users to re-use context mentioned in the conversation and re-phrased by the AI model. Some users found starter phrases to be useful but always edited them to personalize their answer.

8.7 Summary

This final study contributes the design and evaluation of COMPA, a system that explored different contextual grounding strategies to support AAC users and conversation partners in loosening the timing and relevance constraints. A user study including both AAC users and a communication partner of their choice (N=8) indicated that marking the context of a conversation an AAC user wants to speak to and notifying a partner about this context and the intention of the AAC user is useful and helps AAC actively participate in conversation. COMPA's features "leveled the playing field" for AAC users by leveraging opportunities found through our agency framework.

Part IV

**Contributions, Future Work and
Conclusion**

Chapter 9

Contributions, Future Work, and Conclusion

9.1 Contributions

This dissertation demonstrates how social constraints embedded in the traditional ways we communicate can impact an augmented communicator's agency and participation in conversation. By designing technologies that address these social constraints through participatory approaches, we can create new interactions that support user agency. This body of work proposes new approaches in AAC through studies that exemplify how we can design with these constraints in mind to create agency-increasing communication tools. It is my hope that this work can enhance our ability to design interactive AAC systems that are responsive to the shifting nature of conversations and unpredictable partner's behaviors.

9.1.1 Theoretical Contributions

How do ACs exercise conversational agency in the face of constraints such as the AAC device they use, their relationship with the people they are talking with, the task at hand and general social conversational constraints?

This dissertation demonstrates that AAC users are not only impacted by the technology's latency but by social constraints that play a role on how much they can say, how they can say it, and when they can say it. Through looking at participation metrics we can create a deeper understanding of conversational agency that can inform the design of future technologies. Through behavioral coding, a method that turns observed behaviors into metrics, I have quantified turns taken in a conversation, labelled communicative actions taken by different actors, and described how shared decision making happens to analyze how user experience is impacted by social behaviors and technology constraints specific to AAC. Using these metrics, I have contributed Conversational Agency as a new frame to study AAC that describes design parameters for new communication tools. These design parameters or areas of opportunity include finding ways to increase AAC users' control over conversation dynamics and the content of a conversation.

9.1.2 Empirical Contributions

What participation metrics and conversational tasks can help us evaluate agency in aided conversation?

This dissertation also contributes concrete metrics and conversational tasks for measuring agency in AAC. Agency is understood within the context of specific conversational tasks and scenarios with different types of conversation partners. In chapter 3, I explored the differences between technical agency (participating in conversation) and colloquial agency (when participating becomes achieving a specific goal).

Metrics. Using agency as a desired outcome in AAC technology design and presenting metrics that enable researchers to understand how agency is impacted in context, can provide designers with new ways of evaluating and refining empowering interaction techniques and interfaces, in which users experience an instinctive sense of control and ownership over their actions. Metrics contributed in this dissertation include two coding schemes, one detailed in Chapter 3 that details specific AC communicative functions such as: requesting help or time, giving an open-ended explanation, granting/denying permission, and confirming/denying information the CP has said, to measure technical agency.

A second coding scheme is contributed in Chapter 8. This coding scheme includes types of replies and types of conversation initiations that can help researchers better understand how often AAC users use contextual references and details in their responses and how often they open up the conversation to new conversational topics.

Conversational tasks. This dissertation used an interview task to study how communication partners (CPs) supported and AAC user in answering questions for which both the AAC user and conversational partner had knowledge about. A second task that is unique to this dissertation is that of collaborative collage making to understand how AAC users are able to turn their participation or technical agency (expression of their creative ideas) into actual ideas translated to the collage by the CP. To evaluate the last system presented in this dissertation COMPA, we also contribute a conversational task used to simulate a natural conversation where the moderator also takes on the role of asking questions but also committing intentional mistakes such as introducing big topic changes through anecdotes to study how AAC users and CPs maintain a common ground.

9.1.3 Technical and Design Contributions

What systems and design features increase conversational agency for AAC users? What specific technology materials (modalities and platforms) can support conversational agency in AAC?

This dissertation also contributes design knowledge on how different technology materials such as expressive robots and Large Language Models can impact an augmented communicator's agency and their partners behaviors and perceptions. This dissertation demonstrates that using agency as a lens for AAC research can uncover new design opportunities.

The first study that describes Conversational Agency in AAC (Chapter 3) demonstrated how ACs often use a range of nonverbal behaviors (e.g., facial gestures, body orientation) but these are hard to interpret by others. In the field of Human-Robot Interaction (HRI), social robots have used motion to support people in a variety of related communication contexts through nonverbal behavior. Social robot gaze has shown to facilitate turn-taking and robot body gestures have been explored to mediate interpersonal conflict. In Part II of this dissertation, I designed a co-design workshop to identify how social robots could support ACs' nonverbal communication. By using the metaphor of "sidekick", I explore robot designs that support ACs, rather than function as independent agents.

Similarly, a thorough understanding of conversational agency, presented in Chapter 3, demonstrated that in CPs help ACs contextualize their comments by providing explanations on ACs' behalf and interpreted ACs' gestures and vocalizations so they did not have to always rely on their device to respond in time. Nonetheless, when CCPs gave long explanations on behalf of an AC

they often introduced new topics that moved the conversation forward too quickly. By analyzing the rituals among CCPs and ACs we found that future technologies that can retrieve content from a conversation and clarify misunderstandings could support ACs in replying to questions when needed or reusing conversational context without extra effort and without the need of CCP intervention that may in some cases distract from the ACs intention. We explored these opportunities with language technologies in Part III of this thesis which resulted in two new approaches to support AAC users: Speech Macros (chapter 7) and Augmented Contextual Grounding Strategies via the COMPA system (Chapter 8).

9.1.4 Methodological Contributions

What are accessible and inclusive research methods to envision and build new communication technologies with augmented communicators?

This thesis also contributes accessible design methods to envision and investigate potential new technologies with AAC users and their conversation partners. Participatory methods can help us envision new technologies with people whose lived experience is different to our own and calls for the inclusion of relevant stakeholders from different disciplines. In my work, this has involved inviting puppeteers to help us design robot motions (Chapter 4), and data scientists and AI engineers to collaborate with AAC users and imagine the future of language models for communication support (Chapter 6). Conducting participatory research with AAC users can be challenging as they may face accessibility barriers with design materials, difficulty in communicating quickly, and difficulty in collecting data over time [8]. Through co-designing prototypes of future AAC systems I have learned how to design participatory workshops that maximize AAC users participation without causing fatigue.

Conducting participatory research with AAC users can be challenging for many reasons as they may face accessibility barriers with design materials, difficulty in communicating quickly, and difficulty in collecting data over time [8]. The increased obstacles to in-person research created by the COVID-19 pandemic may further impact the ability of AAC users to participate in research. Nonetheless remote participation in research opened doors for AAC users who do not traditionally get a chance to participate. Due to the pandemic all the studies included in Part III of this dissertation were conducted remotely. Here I describe some of the strategies used to support participation from AAC users in remote video sessions developed by myself and my research collaborators:

Combining synchronous and asynchronous communication. In many cases, AAC users may communicate more slowly than the average person, and thus may have difficulty sharing their thoughts during a single research session. In both in-person studies and remote studies, we have found it useful to provide questions to participants before and after the research session, so that they can prepare materials, and a provide feedback using retrospective questionnaires where they can share any thoughts they were unable to share during the session.

Supporting multimodal conversation. During study sessions, we have found it important to allow participants to communicate using multiple modes, including voice chat (using the individual's voice and/or AAC), text chat, or typing in a shared document. This can be especially helpful if there are issues with capturing the user's AAC output in audio, if the AAC is mispronouncing certain words, or if the session includes multiple AAC users or an AAC user and caregivers.

Designing prototypes for screen sharing or web-use. Creating prototypes that are optimized for screen sharing or developing browser extensions like COMPA can help AAC users access prototypes from their device or computers. For screen-shared prototypes, the remote participant can suggest inputs via audio or text chat, which are then entered into the prototype by the tester. The prototypes also support screen resizing to ensure they are readable and can easily adapt to each user's visual acuity and screen setup. While this method is unsuitable for some prototypes, we have found this method to be useful for testing simple applications with minimal setup.

9.1.5 Agency Questions that Remain

Agency in AAC, the ability for AAC users to have control over their communication, as defined by the work in this thesis is impacted by technical and social constraints. In conversation, agency is not only about what the AAC user can do with their device but how much they express, when they express it, and how they express it. As such, conversational agency in AAC is strongly tied with the way AAC users interact with others and their devices during conversations and how these interactions limit or support their personal expression of agency. To measure agency we have used participation metrics such as number of turns taken by an AAC user and their conversation partner, how many turn, words, and types of communicative actions individuals use to achieve different conversational goals (i.e., share an opinion, express a preference, guide a collage making task). As such we have a better understanding of how AAC users can have agency over conversational dynamics and over conversational context.

Agency over Privacy and Self-expression. Nonetheless, it is important to also understand and study agency beyond conversational participation into matters of privacy and self-expression. The study on Speech Macros in Chapter 7 touched on this topic, and uncovered how Large Language Models (LLMs) and future communication systems that need to access sensitive user data may pose a challenge to AAC users' privacy. The rapid advances of AI-based technologies present multiple opportunities to make AAC devices smarter but will potentially require access to more data to achieve personalized suggestions. Future work is needed to investigate how to responsibly explore these algorithms. Additionally, agency as it is tied to self-expression and self-presentation differs from achieving transactional goals. How can we enable self-expression and customization when implementing emerging technologies to AAC devices? The studies with LLMs and Sidekicks completed in this dissertation also suggest further work is needed to ensure users can have more agency over the tone of a message at the language level (word choice) and multimedia level (by adding motion or sound) to fit a user's desired emotional or nuanced framing in a way that allows them to include their personal touch.

Agency and Culture. In order to achieve full participation of people with disabilities in society we have to consider diverse socioeconomic, cultural, and geographical contexts that can help us inform both local and global approaches to technology development. A greater conversation of agency and how it is tied to culture needs to happen. Some AAC user participants whose perspectives are included in this dissertation belong to multi-cultural families. Current AAC systems do not support them in switching between English and their native languages and cultural expressions. This is a major limitation of agency.

Agency through Making As demonstrated in this body of work, we can learn so much about AAC users' values and needs in a way that reveals tacit knowledge through making. AAC user and people with disabilities in general need a more active seat at the table at early stages of technology design. In order to ensure this, we need to continue developing co-design methodologies that can guarantee the participation of people with disabilities in ideation, prototyping and testing. I personally plan to develop inclusive design workshop guidelines and new design research methodologies and validate them with diverse design teams. My goal is for these new methods to support inclusive technology development and help researchers have better strategies to work with under-researched and under-served disabled populations.

9.2 Conclusions

Research has focused on making AAC devices faster to use but this dissertation demonstrates that AAC users are not only impacted by the technology's latency but by social constraints that play a role on how much they can say, how they can say it, and when they can say it. This dissertation draws attention to the importance of considering AAC users agency as a central design parameter for which to optimize for. This dissertation defined and explored what conversational agency looks like in AAC-based interactions outlining how different conversational constraints and constraints specific to AAC affect augmented communicator's agency [157]. The conversational agency framework described specific opportunities to support AAC users in managing social constraints related to attention and context in conversation.

I have studied how motion expressed through robotic objects can help AAC users hold the floor and manage the participation-shift constraints [156]. I continued this work by building a fully functional robotic sidekick with and for one AAC user [159]. We found that using physical sidekicks that augment nonverbal communication, can help AAC users maintain the conversational floor by moving while the user types. Sidekicks can also request other partners to wait just by moving. Sidekicks also supported AAC users by drawing attention to the user and signaling through movement when the AAC user wanted to take a turn.

While the robotic sidekicks helped us address social constraints related to managing conversational dynamics, we then focused on exploring how language technologies such as Large Language Models could support AAC users in addressing the relevance and ritual constraints in conversation related to the content of conversations. In order to center AC's voices in how language technologies may support them in having conversations in real time I carried out an interdisciplinary participatory workshop, tested AI-based Large language models with ACs and created COMPA a system to augment common ground.

In summary, this dissertation has implications to improve participation of people with disabilities in social conversation and in the design process by (1) expanding on social science theory to produce agency-centered frameworks, (2) innovating on accessible co-design practices, and (3) generating design knowledge on how specific technology materials can create more accessible communication experiences.

Bibliography

- [1] Lucy Anderson-Bashan, Benny Megidish, Hadas Erel, Iddo Wald, Guy Hoffman, Oren Zuckerman, and Andrey Grishko. The greeting machine: an abstract robotic object for opening encounters. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 595–602. IEEE, 2018.
- [2] Sean Andrist, Xiang Zhi Tan, Michael Gleicher, and Bilge Mutlu. Conversational gaze aversion for humanlike robots. In *Proceedings of the 10th International Conference on Human-Robot Interaction (HRI)*. ACM, 2014.
- [3] American Speech Hearing Association ASHA. Augmentative and alternative communication. retrieved, 2021.
- [4] Christine E Ashby. Whose voice is it anyway?: Giving voice and qualitative research involving individuals that type to communicate. *Disability Studies Quarterly*, 31(4), 2011.
- [5] Salena Babb, Jessica Gormley, David McNaughton, and Janice Light. Enhancing independent participation within vocational activities for an adolescent with asd using aac video visual scene displays. *Journal of Special Education Technology*, page 0162643418795842, 2018.
- [6] BR Baker. Minspeak, a semantic compaction system that makes self-expression easier for communicatively disabled individuals. *Byte*, 7(9):186–202, 1982.
- [7] Dare A Baldwin. Understanding the link between joint attention and language. *Joint attention: Its origins and role in development*, pages 131–158, 1995.
- [8] Erin Beneteau. Who are you asking?: Qualitative methods for involving aac users as primary research participants. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.
- [9] Cynthia L Bennett, Erin Brady, and Stacy M Branham. Interdependence as a frame for assistive technology research and design. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 161–173. ACM, 2018.
- [10] Cynthia L Bennett, Burren Peil, and Daniela K Rosner. Biographical prototypes: Reimagining recognition and disability in design. In *Proceedings of the 2019 on Designing Interactive Systems Conference*, pages 35–47. ACM, 2019.
- [11] Cynthia L Bennett, Kristen Shinohara, Brianna Blaser, Andrew Davidson, and Kat M Steele. Using a design workshop to explore accessible ideation. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 303–304. ACM, 2016.
- [12] David Beukelman and Janice Light. *Augmentative and alternative communication: supporting children and adults with complex communication needs*. Paul H. Brookes Publishing Co, Baltimore, Maryland, 2020.

- [13] David R Beukelman and Pat Mirenda. *Augmentative and alternative communication: Supporting children and adults with complex communication needs*. Paul H. Brookes Pub., 2013.
- [14] David R Beukelman, Pat Mirenda, et al. *Augmentative and alternative communication*. Paul H. Brookes Baltimore, 1998.
- [15] Sarah W Blackstone. Circles of communication partners. *Augmentative Communication News*, 12:1–2, 1999.
- [16] Sarah W. Blackstone and Barbara Collier. Communication assistants: Human supports for communication access. *Augmentative Communication News*, 20(3):1–15, 2008.
- [17] Sarah W Blackstone, Michael B Williams, and David P Wilkins. Key principles underlying research and practice in aac. *Augmentative and alternative communication*, 23(3):191–203, 2007.
- [18] Isaac Blankensmith and Google. Paper signals, 2021.
- [19] Gautam Bose, Marisa Lu, Lucas Ochoa, and Dan Lockton. Emoto: From phone to emotive robotic AI sidekick. *TEI 2019 - Proceedings of the 13th International Conference on Tangible, Embedded, and Embodied Interaction*, pages 451–457, 2019.
- [20] Danielle Bragg, Oscar Koller, Mary Bellard, Larwan Berke, Patrick Boudreault, Annelies Braffort, Naomi Caselli, Matt Huenerfauth, Hernisa Kacorri, Tessa Verhoef, et al. Sign language recognition, generation, and translation: An interdisciplinary perspective. In *The 21st international ACM SIGACCESS conference on computers and accessibility*, pages 16–31, 2019.
- [21] Eva Brandt, Thomas Binder, and Elizabeth BN Sanders. Ways to engage telling, making and enacting. *Routledge international handbook of participatory design*. Routledge, New York, pages 145–181, 2012.
- [22] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.
- [23] Robin N. Brewer. Facilitating discussion and shared meaning: Rethinking co-design sessions with people with vision impairments. In *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '18, page 258–262, New York, NY, USA, 2018. Association for Computing Machinery.
- [24] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc., 2020.
- [25] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

- [26] Diane Nelson Bryen. Communication during times of natural or man-made emergencies. *Journal of Pediatric Rehabilitation Medicine*, 2(2):123–129, 2009.
- [27] Shanqing Cai, Subhashini Venugopalan, Katrin Tomanek, Ajit Narayanan, Meredith R Morris, and Michael P Brenner. Context-aware abbreviation expansion using large language models. *arXiv preprint arXiv:2205.03767*, 2022.
- [28] Jessica R Cauchard, Kevin Y Zhai, Marco Spadafora, and James A Landay. Emotion encoding in human-drone interaction. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 263–270. IEEE, 2016.
- [29] Elizabeth Clark, Tal August, Sofia Serrano, Nikita Haduong, Suchin Gururangan, and Noah A Smith. All that’s ‘human’ is not gold: Evaluating human evaluation of generated text. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7282–7296, Online, August 2021. Association for Computational Linguistics.
- [30] Herbert H Clark, Susan E Brennan, et al. Grounding in communication. *Perspectives on socially shared cognition*, 13(1991):127–149, 1991.
- [31] Michael Clarke and Andrea Kirton. Patterns of interaction between children with physical disabilities using augmentative and alternative communication systems and their peers. *Child Language Teaching and Therapy*, 19(2):135–151, 2003.
- [32] Michael Clarke and Ray Wilkinson. Interaction between children with cerebral palsy and their peers 2: Understanding initiated voca-mediated turns. *Augmentative and Alternative Communication*, 24(1):3–15, 2008.
- [33] David Coyle, James Moore, Per Ola Kristensson, Paul Fletcher, and Alan Blackwell. I did that! measuring users’ experience of agency in their own actions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2025–2034. ACM, 2012.
- [34] Humphrey Curtis, Timothy Neate, and Carlota Vazquez Gonzalez. State of the art in aac: A systematic review and taxonomy. *State of the Art in AAC: A Systematic Review and Taxonomy*, 2022.
- [35] Jiamin Dai, Karyn Moffatt, Jinglan Lin, and Khai Truong. Designing for relational maintenance: New directions for aac research. In *CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2022.
- [36] Adriane B Davis, Melody M Moore, and Veda C Storey. Context-aware communication for severely disabled users. In *Proceedings of the 2003 conference on Universal usability*, pages 106–111, 2002.
- [37] Lieven De Couvreur and Richard Goossens. Design for (every) one: co-creation as a bridge between universal design and rehabilitation engineering. *CoDesign*, 7(2):107–121, 2011.
- [38] Katherine Deibel. A convenient heuristic model for understanding assistive technology adoption. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–2, 2013.
- [39] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019*

- Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- [40] Kirsten Ellis, Ross De Vent, Reuben Kirkham, and Patrick Olivier. Bespoke reflections: Creating a one-handed braille keyboard. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–13, 2020.
- [41] Mustafa Emirbayer and Ann Mische. What is agency? *American journal of sociology*, 103(4):962–1023, 1998.
- [42] Nicholas James Enfield and Paul Kockelman. *Distributed agency*. Oxford University Press, 2017.
- [43] Nick J Enfield. *Relationship thinking: Agency, enchrony, and human sociality*. Oxford University Press, 2013.
- [44] Christopher R Engelke and D Jeffery Higginbotham. Looking to speak: On the temporality of misalignment in interaction involving an augmented communicator using eye-gaze technology. *Journal of Interactional Research in Communication Disorders*, 4(1):95–122, 2013.
- [45] Alexander Fiannaca, Jon Campbell, Ann Paradiso, and Meredith Ringel Morris. Voice-setting: Voice Authoring UIs for Improved Expressivity in Augmentative Communication. 2018.
- [46] Alexander Fiannaca, Ann Paradiso, Mira Shah, and Meredith Ringel Morris. Aacrobat: Using mobile devices to lower communication barriers and provide autonomy with gaze-based aac. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 683–695. ACM, 2017.
- [47] Mauricio Fontana de Vargas, Jiamin Dai, and Karyn Moffatt. Aac with automated vocabulary from photographs: Insights from school and speech-language therapy settings. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '22, New York, NY, USA, 2022. Association for Computing Machinery.
- [48] Mauricio Fontana de Vargas and Karyn Moffatt. Automated generation of storytelling vocabulary from photographs for use in AAC. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1353–1364, Online, August 2021. Association for Computational Linguistics.
- [49] M Fried-Oken, E Jakobs, and T Jakobs. Smart predict: Aac app that integrates partner knowledge into word prediction. In *annual Assistive Technology Industry Association (ATIA) conference Orlando, FL.[Google Scholar]*, 2018.
- [50] Melanie Fried-Oken, Charity Rowland, Glory Baker, Mayling Dixon, Carolyn Mills, Darlene Schultz, and Barry Oken. The effect of voice output on aac-supported conversations of persons with alzheimer’s disease. *ACM Transactions on Accessible Computing (TACCESS)*, 1(3):1–11, 2009.
- [51] Batya Friedman and David G Hendry. *Value sensitive design: Shaping technology with moral imagination*. Mit Press, 2019.

- [52] Julia Galliers, Stephanie Wilson, Abi Roper, Naomi Cocks, Jane Marshall, Sam Muscroft, and Tim Pring. Words are not enough: Empowering people with aphasia in the design process. *ACM International Conference Proceeding Series*, 1:51–60, 2012.
- [53] Katy Ilonka Gero and Lydia B Chilton. Metaphoria: An algorithmic companion for metaphor creation. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–12, 2019.
- [54] David R Gibson. Seizing the moment: The problem of conversational agency. *Sociological Theory*, 18(3):368–382, 2000.
- [55] Erving Goffman et al. *The presentation of self in everyday life*. Harmondsworth London, 1978.
- [56] Steven M Goodman, Erin Buehler, Patrick Clary, Andy Coenen, Aaron Donsbach, Tiffanie N Horne, Michal Lahav, Robert Macdonald, Rain Breaw Michaels, Ajit Narayanan, et al. Lampost: Design and evaluation of an ai-assisted email writing prototype for adults with dyslexia. *arXiv preprint arXiv:2207.02308*, 2022.
- [57] Charles Goodwin. Contextures of action. *Embodied interaction: Language and body in the material world*, pages 182–193, 2011.
- [58] Charles Goodwin. The co-operative, transformative organization of human action and knowledge. *Journal of pragmatics*, 46(1):8–23, 2013.
- [59] Carla F Griggio, Joanna Mcgrener, and Wendy E Mackay. Customizations and expression breakdowns in ecosystems of communication apps. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–26, 2019.
- [60] Bruce Hanington and Bella Martin. *Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions*. Rockport Publishers, 2012.
- [61] Christina Harrington, Sheena Erete, and Anne Marie Piper. Deconstructing community-based collaborative design: Towards more equitable participatory design engagements. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–25, 2019.
- [62] D Jeffery Higginbotham. In-person interaction in aac: New perspectives on utterances, multimodality, timing, and device design. *Perspectives on Augmentative and Alternative Communication*, 18(4):154–160, 2009.
- [63] D Jeffery Higginbotham, Ann M Bisantz, Michelle Sunm, Kim Adams, and Fen Yik. The effect of context priming and task type on augmentative communication performance. *Augmentative and Alternative Communication*, 25(1):19–31, 2009.
- [64] D Jeffery Higginbotham, Gregory W Lesh, Bryan J Moulton, and Brian Roark. The application of natural language processing to augmentative and alternative communication. *Assistive Technology*, 24(1):14–24, 2012.
- [65] D Jeffery Higginbotham and DP Wilkins. Slipping through the timestream: Social issues of time and timing in augmented interactions. *Constructing (in) competence: Disabling evaluations in clinical and social interaction*, 2:49–82, 1999.
- [66] DJ Higginbotham, K Fulcher, and J Seale. Time and timing in interactions involving individuals with als, their unimpaired partners and their speech generating devices. *The silent partner*, pages 199–229, 2016.

- [67] Clare Hocking. Function or feelings: factors in abandonment of assistive devices. *Technology and Disability*, 11(1-2):3–11, 1999.
- [68] Guy Hoffman, Oren Zuckerman, Gilad Hirschberger, Michal Luria, and Tal Shani-Sherman. Design and evaluation of a peripheral robotic conversation companion. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 3–10. IEEE, 2015.
- [69] Amy Hurst and Jasmine Tobias. Empowering individuals with do-it-yourself assistive technology. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*, pages 11–18, 2011.
- [70] Seray B Ibrahim, Asimina Vasalou, and Michael Clarke. Design opportunities for aac and children with severe speech and physical impairments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 227. ACM, 2018.
- [71] Ryo Iijima, Akihisa Shitara, Sayan Sarcar, and Yoichi Ochiai. Word cloud for meeting: A visualization system for dhh people in online meetings. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–4, 2021.
- [72] Hiroshi Ishii. Tangible bits: beyond pixels. In *Proceedings of the 2nd international conference on Tangible and embedded interaction*, pages xv–xxv, 2008.
- [73] Shervin Javdani, Henny Admoni, Stefania Pellegrinelli, Siddhartha S. Srinivasa, and J. Andrew Bagnell. Shared autonomy via hindsight optimization for teleoperation and collaboration. *International Journal of Robotics Research*, 37:717–742, 2018.
- [74] Ellen Jiang, Kristen Olson, Edwin Toh, Alejandra Molina, Aaron Donsbach, Michael Terry, and Carrie J Cai. Promptmaker: Prompt-based prototyping with large language models. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems, CHI EA '22*, New York, NY, USA, 2022. Association for Computing Machinery.
- [75] Dan Jurafsky, Liz Shriberg, and Debra Biasca. Ws-97 switchboard damsl coders manual, 1997.
- [76] Shaun K Kane, Barbara Linam-Church, Kyle Althoff, and Denise McCall. What we talk about: designing a context-aware communication tool for people with aphasia. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*, pages 49–56, 2012.
- [77] Shaun K. Kane, Meredith Ringel Morris, Ann Paradiso, and Jon Campbell. At times avuncular and cantankerous, with the reflexes of a mongoose: Understanding self-expression through augmentative and alternative communication devices. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW '17*, pages 1166–1179, New York, NY, USA, 2017. ACM.
- [78] S.K. Kane and M.R. Morris. Let’s Talk about X: Combining image recognition and eye gaze to support conversation for people with ALS. *DIS 2017 - Proceedings of the 2017 ACM Conference on Designing Interactive Systems*, pages 129–134, 2017.
- [79] Ryan Kelly, Daniel Gooch, Bhagyashree Patil, and Leon Watts. Demanding by design: Supporting effortful communication practices in close personal relationships. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 70–83, 2017.

- [80] Nitish Shirish Keskar, Bryan McCann, Lav R Varshney, Caiming Xiong, and Richard Socher. Ctrl: A conditional transformer language model for controllable generation. *arXiv preprint arXiv:1909.05858*, 2019.
- [81] Mark L Knapp, Judith A Hall, and Terrence G Horgan. *Nonverbal communication in human interaction*. Cengage Learning, 2013.
- [82] Heather Knight, Timothy Lee, Brittany Hallawell, and Wendy Ju. I get it already! the influence of chairbot motion gestures on bystander response. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 443–448. IEEE, 2017.
- [83] Spyros Kousidis and David Schlangen. The power of a glance: Evaluating embodiment and turn-tracking strategies of an active robotic overhearer. In *2015 AAI Spring Symposium Series*, 2015.
- [84] Per Ola Kristensson, James Lilley, Rolf Black, and Annalu Waller. A design engineering approach for quantitatively exploring context-aware sentence retrieval for nonspeaking individuals with motor disabilities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2020.
- [85] Kundan Kumar, Rithesh Kumar, Thibault de Boissiere, Lucas Gestin, Wei Zhen Teoh, Jose Sotelo, Alexandre de Brébisson, Yoshua Bengio, and Aaron C Courville. Melgan: Generative adversarial networks for conditional waveform synthesis. *Advances in neural information processing systems*, 32, 2019.
- [86] Amanda Lazar, Caroline Edasis, and Anne Marie Piper. A critical lens on dementia and design in hci. In *Chi*, pages 2175–2188, 2017.
- [87] Amanda Lazar, Jessica L Feuston, Caroline Edasis, and Anne Marie Piper. Making as expression: Informing design with people with complex communication needs through art therapy. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 351. ACM, 2018.
- [88] Kenton Lee, Luheng He, Mike Lewis, and Luke Zettlemoyer. End-to-end neural coreference resolution. *arXiv preprint arXiv:1707.07045*, 2017.
- [89] Gregory W Lesher and D Jeffery Higginbotham. Using web content to enhance augmentative communication. In *Proceedings of CSUN*, volume 2005, 2005.
- [90] Janice Light. Toward a definition of communicative competence for individuals using augmentative and alternative communication systems. *Augmentative and alternative communication*, 5(2):137–144, 1989.
- [91] Janice Light, Barbara Collier, and Penny Parnes. Communicative interaction between young nonspeaking physically disabled children and their primary caregivers: Part ii-communicative function. *Augmentative and Alternative Communication*, 1(3):98–107, 1985.
- [92] Janice Light and David McNaughton. Communicative competence for individuals who require augmentative and alternative communication: A new definition for a new era of communication?, 2014.
- [93] Janice Light, David McNaughton, David Beukelman, Susan Koch Fager, Melanie Fried-Oken, Thomas Jakobs, and Erik Jakobs. Challenges and opportunities in augmentative and

- alternative communication: Research and technology development to enhance communication and participation for individuals with complex communication needs. *Augmentative and Alternative Communication*, 35(1):1–12, 2019.
- [94] Hannah Limerick, David Coyle, and James W Moore. The experience of agency in human-computer interactions: a review. *Frontiers in human neuroscience*, 8:643, 2014.
- [95] Dan Lockton, Devika Singh, Saloni Sabnis, Michelle Chou, Sarah Foley, and Alejandro Pantoja. New metaphors: A workshop method for generating ideas and reframing problems in design and beyond. In *Proceedings of the 2019 on Creativity and Cognition*, pages 319–332. ACM, 2019.
- [96] John Long. *Approaches and Frameworks for HCI Research*. Cambridge University Press, 2021.
- [97] Michal Luria. Designing robot personality based on fictional sidekick characters. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 307–308. ACM, 2018.
- [98] Michal Luria, Guy Hoffman, Benny Megidish, Oren Zuckerman, and Sung Park. Designing vyo, a robotic smart home assistant: Bridging the gap between device and social agent. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 1019–1025. IEEE, 2016.
- [99] Kelly Mack, Emma McDonnell, Dhruv Jain, Lucy Lu Wang, Jon E. Froehlich, and Leah Findlater. What do we mean by “accessibility research”? a literature survey of accessibility papers in chi and assets from 1994 to 2019. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2021.
- [100] Jennifer Mankoff, Gillian R Hayes, and Devva Kasnitz. Disability studies as a source of critical inquiry for the field of assistive technology. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*, pages 3–10, 2010.
- [101] Yoichi Matsuyama, Iwao Akiba, Shinya Fujie, and Tetsunori Kobayashi. Four-participant group conversation: A facilitation robot controlling engagement density as the fourth participant. *Computer Speech & Language*, 33(1):1–24, 2015.
- [102] Michelle Mckelvey, David L Evans, Norimune Kawai, and David Beukelman. Communication styles of persons with als as recounted by surviving partners. *Augmentative and Alternative Communication*, 28(4):232–242, 2012.
- [103] Matthew B Miles, A Michael Huberman, Michael A Huberman, and Michael Huberman. *Qualitative data analysis: An expanded sourcebook*. sage, 1994.
- [104] Timothy Mills, H Timothy Bunnell, and Rupal Patel. Towards personalized speech synthesis for augmentative and alternative communication. *Augmentative and Alternative Communication*, 30(3):226–236, 2014.
- [105] Sanika Moharana, Alejandro E Panduro, Hee Rin Lee, and Laurel D Riek. Robots for joy, robots for sorrow: community based robot design for dementia caregivers. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 458–467. IEEE, 2019.
- [106] Bilge Mutlu, Toshiyuki Shiwa, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 61–68. ACM, 2009.

- [107] Yukiko I Nakano, Takashi Yoshino, Misato Yatsushiro, and Yutaka Takase. Generating robot gaze on the basis of participation roles and dominance estimation in multiparty interaction. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 5(4):22, 2016.
- [108] Tong Niu and Mohit Bansal. Polite dialogue generation without parallel data. *Transactions of the Association for Computational Linguistics*, 6:373–389, 2018.
- [109] Don Norman. *The design of everyday things: Revised and expanded edition*. Basic books, 2013.
- [110] OpenAI. Chatgpt large language model (gpt-3.5-turbo), 2023.
- [111] Joohee Park, Young Woo Park, and Tek Jin Nam. Wrigglo: Shape-changing peripheral for interpersonal mobile communication. *Conference on Human Factors in Computing Systems - Proceedings*, pages 599–602, 2014.
- [112] Rupal Patel and Rajiv Radhakrishnan. Enhancing access to situational vocabulary by leveraging geographic context. *Assistive Technology Outcomes and Benefits*, 4(1):99–114, 2007.
- [113] Owain Pedgley. Capturing and analysing own design activity. *Design studies*, 28(5):463–483, 2007.
- [114] Haoruo Peng, Yangqiu Song, and Dan Roth. Event detection and co-reference with minimal supervision. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 392–402, 2016.
- [115] E Margaret Perkoff. Dialogue act analysis for alternative and augmentative communication. 2021.
- [116] Betsy Phillips and Hongxin Zhao. Predictors of assistive technology abandonment. *Assistive technology*, 5(1):36–45, 1993.
- [117] Wei Ping, Kainan Peng, and Jitong Chen. Clarinet: Parallel wave generation in end-to-end text-to-speech. *arXiv preprint arXiv:1807.07281*, 2018.
- [118] Graham Pullin and Shannon Hennig. 17 ways to say yes: Toward nuanced tone of voice in aac and speech technology. *Augmentative and Alternative Communication*, 31(2):170–180, 2015.
- [119] Graham Pullin and Shannon Hennig. 17 ways to say yes: Toward nuanced tone of voice in aac and speech technology. *Augmentative and Alternative Communication*, 31(2):170–180, 2015.
- [120] Graham Pullin and Alan Newell. Focussing on extra-ordinary users. In *International Conference on Universal Access in Human-Computer Interaction*, pages 253–262. Springer, 2007.
- [121] Graham Pullin, Jutta Treviranus, Rupal Patel, and Jeff Higginbotham. Designing interaction, voice, and inclusion in aac research. *Augmentative and Alternative Communication*, 33(3):139–148, 2017.
- [122] Bodil Ravneberg. Usability and abandonment of assistive technology. *Journal of Assistive Technologies*, 2012.
- [123] D Raj Reddy. Speech recognition by machine: A review. *Proceedings of the IEEE*, 64(4):501–531, 1976.

- [124] Brian Roark, Melanie Fried-Oken, and Chris Gibbons. Huffman and linear scanning methods with statistical language models. *Augmentative and Alternative Communication*, 31(1):37–50, 2015.
- [125] Albert B Robillard. Communication problems in the intensive care unit. *Qualitative Sociology*, 17(4):383–395, 1994.
- [126] Ben Robins, Kerstin Dautenhahn, and Paul Dickerson. From isolation to communication: a case study evaluation of robot assisted play for children with autism with a minimally expressive humanoid robot. In *2009 Second International Conferences on Advances in Computer-Human Interactions*, pages 205–211. IEEE, 2009.
- [127] Harvey Sacks, Emanuel A Schegloff, and Gail Jefferson. A simplest systematics for the organization of turn taking for conversation. In *Studies in the organization of conversational interaction*, pages 7–55. Elsevier, 1978.
- [128] Elizabeth B-N Sanders and Pieter Jan Stappers. Co-creation and the new landscapes of design. *Co-design*, 4(1):5–18, 2008.
- [129] Elizabeth B-N Sanders and Pieter Jan Stappers. *Convivial design toolbox*. 2012.
- [130] Antara Satchidanand, Jeff Higginbotham, Ann Bisantz, Naif Aldhaam, Ahmed Elsayed, Iman Carr, Ahmed A Hussein, and Khurshid Guru. “put the what, where? cut here?!” challenges to coordinating attention in robot-assisted surgery: a microanalytic pilot study. *BMJ open*, 11(7):e046132, 2021.
- [131] Irina Savolainen, Anu Klippi, Tuula Tykkyläinen, Jeff Higginbotham, and Kaisa Launonen. The structure of participants’ turn-transition practices in aided conversations that use speech-output technologies. *Augmentative and alternative communication*, 36(1):18–30, 2020.
- [132] Igor Schadle, Jean-Yves Antoine, Brigitte Le Pévédic, and Franck Poirier. Sibyl – aac system using nlp techniques. 07 2004.
- [133] Emanuel A Schegloff. Practices and actions: Boundary cases of other-initiated repair. *Discourse processes*, 23(3):499–545, 1997.
- [134] Dennis Schleicher, Peter Jones, and Oksana Kachur. Bodystorming as embodied designing. *interactions*, 17(6):47–51, 2010.
- [135] Donald A Schön. Designing as reflective conversation with the materials of a design situation. *Knowledge-based systems*, 5(1):3–14, 1992.
- [136] Jennifer M Seale, Ann M Bisantz, and Jeff Higginbotham. Interaction symmetry: Assessing augmented speaker and oral speaker performances across four tasks. *Augmentative and Alternative Communication*, 36(2):82–94, 2020.
- [137] Samuel C Sennott, Linda Akagi, Mary Lee, and Anthony Rhodes. Aac and artificial intelligence (ai). *Topics in language disorders*, 39(4):389, 2019.
- [138] Junxiao Shen, Boyin Yang, John J Dudley, and Per Ola Kristensson. Kwickchat: A multi-turn dialogue system for aac using context-aware sentence generation by bag-of-keywords. In *27th International Conference on Intelligent User Interfaces, IUI ’22*, page 853–867, New York, NY, USA, 2022. Association for Computing Machinery.
- [139] Kristen Shinohara and Jacob O Wobbrock. In the shadow of misperception: assistive technology use and social interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 705–714. ACM, 2011.

- [140] Kristen Shinohara, Jacob O. Wobbrock, and Wanda Pratt. Incorporating social factors in accessible design. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '18, page 149–160, New York, NY, USA, 2018. Association for Computing Machinery.
- [141] Ben Shneiderman, Catherine Plaisant, Maxine S Cohen, Steven Jacobs, Niklas Elmqvist, and Nicholas Diakopoulos. *Designing the user interface: strategies for effective human-computer interaction*. Pearson, 2016.
- [142] David Sirkin and Wendy Ju. Using embodied design improvisation as a design research tool. In *Proceedings of the international conference on Human Behavior in Design (HBiD 2014)*, Ascona, Switzerland, 2014.
- [143] David Sirkin, Brian Mok, Stephen Yang, and Wendy Ju. Mechanical ottoman: how robotic furniture offers and withdraws support. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 11–18, 2015.
- [144] David Sirkin, Gina Venolia, John Tang, George Robertson, Taemie Kim, Kori Inkpen, Mara Sedlins, Bongshin Lee, and Mike Sinclair. Motion and attention in a kinetic videoconferencing proxy. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6946 LNCS(PART 1):162–180, 2011.
- [145] Mona Sloane, Emanuel Moss, Olaitan Awomolo, and Laura Forlano. Participation is not a design fix for machine learning. *arXiv preprint arXiv:2007.02423*, 2020.
- [146] Kiley Sobel, Alexander Fiannaca, Jon Campbell, Harish Kulkarni, Ann Paradiso, Ed Cutrell, and Meredith Ringel Morris. Exploring the design space of aac awareness displays. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 2890–2903. ACM, 2017.
- [147] Gabriele B Sweidel. Management strategies in the communication of speaking persons and persons with a speech disability. *Research on Language & Social Interaction*, 25(1-4):195–214, 1991.
- [148] Xu Tan, Tao Qin, Frank Soong, and Tie-Yan Liu. A survey on neural speech synthesis. *arXiv preprint arXiv:2106.15561*, 2021.
- [149] Paul Taylor and Amy Isard. Ssml: A speech synthesis markup language. *Speech communication*, 21(1-2):123–133, 1997.
- [150] Hamish Tennent, Solace Shen, and Malte Jung. Micbot: A peripheral robotic object to shape conversational dynamics and team performance. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 133–142. IEEE, 2019.
- [151] Romal Thoppilan, Daniel De Freitas, Jamie Hall, Noam Shazeer, Apoorv Kulshreshtha, Heng-Tze Cheng, Alicia Jin, Taylor Bos, Leslie Baker, Yu Du, YaGuang Li, Hongrae Lee, Huaixiu Steven Zheng, Amin Ghafouri, Marcelo Menegali, Yanping Huang, Maxim Krikun, Dmitry Lepikhin, James Qin, Dehao Chen, Yuanzhong Xu, Zhifeng Chen, Adam Roberts, Maarten Bosma, Vincent Zhao, Yanqi Zhou, Chung-Ching Chang, Igor Krivokon, Will Rusch, Marc Pickett, Pranesh Srinivasan, Laichee Man, Kathleen Meier-Hellstern, Meredith Ringel Morris, Tulsee Doshi, Renelito Delos Santos, Toju Duke, Johnny Soraker, Ben Zevenbergen, Vinodkumar Prabhakaran, Mark Diaz, Ben Hutchinson, Kristen Olson, Alejandra Molina, Erin Hoffman-John, Josh Lee, Lora Aroyo, Ravi Rajakumar, Alena Butryna, Matthew Lamm, Viktoriya Kuzmina, Joe Fenton, Aaron Cohen, Rachel

- Bernstein, Ray Kurzweil, Blaise Aguera-Arcas, Claire Cui, Marian Croak, Ed Chi, and Quoc Le. *Lamda: Language models for dialog applications*, 2022.
- [152] John Todman. Rate and quality of conversations using a text-storage aac system: Single-case training study. *Augmentative and Alternative Communication*, 16(3):164–179, 2000.
- [153] John Todman, David Rankin, and Portia File. The use of stored text in computer-aided conversation: A single-case experiment. *Journal of language and social psychology*, 18(3):287–309, 1999.
- [154] Keith Trnka, John McCaw, Debra Yarrington, Kathleen F McCoy, and Christopher Pennington. User interaction with word prediction: The effects of prediction quality. *ACM Transactions on Accessible Computing (TACCESS)*, 1(3):1–34, 2009.
- [155] Khai N Truong, Gillian R Hayes, and Gregory D Abowd. Storyboarding: an empirical determination of best practices and effective guidelines. In *Proceedings of the 6th conference on Designing Interactive systems*, pages 12–21, 2006.
- [156] Stephanie Valencia, Michal Luria, Amy Pavel, Jeffrey P. Bigham, and Henny Admoni. Co-designing socially assistive sidekicks for motion-based aac. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, HRI '21*, page 24–33, New York, NY, USA, 2021. Association for Computing Machinery.
- [157] Stephanie Valencia, Amy Pavel, Jared Santa Maria, Seunga Yu, Jeffrey P Bigham, and Henny Admoni. Conversational agency in augmentative and alternative communication. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2020.
- [158] Stephanie Valencia, Amy Pavel, Jared Santa Maria, Seunga (Gloria) Yu, Jeffrey P. Bigham, and Henny Admoni. Conversational agency in augmentative and alternative communication. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20*, page 1–12, New York, NY, USA, 2020. Association for Computing Machinery.
- [159] Stephanie Valencia, Mark Steidl, Michael Rivera, Cynthia Bennett, Jeffrey P. Bigham, and Henny Admoni. Aided nonverbal communication through physical expressive objects. In *To appear in proceedings of the 2021 International ACM/SIGACCESS Conference on Computers and Accessibility, ASSETS '21*, New York, NY, USA, 2021. Association for Computing Machinery.
- [160] Keith Vertanen. Towards improving predictive aac using crowdsourced dialogues and partner context. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 347–348. ACM, 2017.
- [161] Annalu Waller. Telling tales: unlocking the potential of aac technologies. *International journal of language & communication disorders*, 54(2):159–169, 2019.
- [162] Joseph B Walther. Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication research*, 23(1):3–43, 1996.
- [163] Yuxuan Wang, Daisy Stanton, Yu Zhang, RJ-Skerry Ryan, Eric Battenberg, Joel Shor, Ying Xiao, Ye Jia, Fei Ren, and Rif A Saurous. Style tokens: Unsupervised style modeling, control and transfer in end-to-end speech synthesis. In *International Conference on Machine Learning*, pages 5180–5189. PMLR, 2018.
- [164] Jacqueline Kory Westlund, Jin Joo Lee, Luke Plummer, Fardad Faridi, Jesse Gray, Matt Berlin, Harald Quintus-Bosz, Robert Hartmann, Mike Hess, Stacy Dyer, et al. Tega: a social

- robot. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 561–561. IEEE, 2016.
- [165] Mikael Wiberg. Methodology for materiality: interaction design research through a material lens. *Personal and ubiquitous computing*, 18(3):625–636, 2014.
- [166] Bruce Wisenburn and D Jeffery Higginbotham. An aac application using speaking partner speech recognition to automatically produce contextually relevant utterances: Objective results. *Augmentative and Alternative Communication*, 24(2):100–109, 2008.
- [167] Bruce Wisenburn and D Jeffery Higginbotham. Participant evaluations of rate and communication efficacy of an aac application using natural language processing. *Augmentative and Alternative Communication*, 25(2):78–89, 2009.
- [168] Jacob O Wobbrock, Shaun K Kane, Krzysztof Z Gajos, Susumu Harada, and Jon Froehlich. Ability-based design: Concept, principles and examples. *ACM Transactions on Accessible Computing (TACCESS)*, 3(3):1–27, 2011.
- [169] Tongshuang Wu, Michael Terry, and Carrie J. Cai. Ai chains: Transparent and controllable human-ai interaction by chaining large language model prompts, 2021.
- [170] Qian Yang. Machine learning as a ux design material: how can we imagine beyond automation, recommenders, and reminders? In *2018 AAAI spring symposium series*, 2018.
- [171] Hao Zhou, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. Emotional chatting machine: Emotional conversation generation with internal and external memory. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.
- [172] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. Research through design as a method for interaction design research in hci. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 493–502. ACM, 2007.
- [173] John Zimmerman, Erik Stolterman, and Jodi Forlizzi. An analysis and critique of research through design: towards a formalization of a research approach. In *proceedings of the 8th ACM conference on designing interactive systems*, pages 310–319, 2010.
- [174] Oren Zuckerman, Dina Walker, Andrey Grishko, Tal Moran, Chen Levy, Barak Lisak, Iddo Yehoshua Wald, and Hadas Erel. Companionship Is Not a Function: The Effect of a Novel Robotic Object on Healthy Older Adults’ Feelings of “Being-Seen”. *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI ’20)*, pages 1–14, 2020.