

Consideration of artificial intelligence applications for interpreting communicative movements by individuals with visual and/or motor disabilities

Krista M. Wilkinson, Dawn J. Sowers, Lynn R. Elko, Mitchell Case, Sharon Redmon, Kevin Williams, Tanuj Namdeo & Syed Billah

To cite this article: Krista M. Wilkinson, Dawn J. Sowers, Lynn R. Elko, Mitchell Case, Sharon Redmon, Kevin Williams, Tanuj Namdeo & Syed Billah (15 May 2025): Consideration of artificial intelligence applications for interpreting communicative movements by individuals with visual and/or motor disabilities, *Augmentative and Alternative Communication*, DOI: [10.1080/07434618.2025.2495905](https://doi.org/10.1080/07434618.2025.2495905)

To link to this article: <https://doi.org/10.1080/07434618.2025.2495905>



© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 15 May 2025.



[Submit your article to this journal](#)



Article views: 681



[View related articles](#)



[View Crossmark data](#)



Citing articles: 2 [View citing articles](#)

Consideration of artificial intelligence applications for interpreting communicative movements by individuals with visual and/or motor disabilities

Krista M. Wilkinson^a, Dawn J. Sowers^a, Lynn R. Elko^b, Mitchell Case^c, Sharon Redmon^d, Kevin Williams^e, Tanuj Namdeo^f and Syed Billah^f

^aDepartment of Communication Sciences & Disorders, the Pennsylvania State University, University Park, PA, USA; ^bSelf Affiliated, Tamaqua, PA, USA; ^cThe Pennsylvania State University, University Park, PA, USA; ^dDepartment of Educational and School Psychology and Special Education, the Pennsylvania State University, University Park, PA, USA; ^eFreelance Contractor/Speaker/Advisor, NC, USA; ^fCollege of Information Sciences and Technology, the Pennsylvania State University, University Park, PA, USA

ABSTRACT

Access to speech-language therapies that promote optimal communication outcomes is a fundamental right outlined by the United Nation's Article 19 of the Convention on the Rights of Persons with Disabilities as well as the Communication Bill of Rights of the National Joint Committee on the Communication Needs of Persons with Severe Disabilities. For many individuals, augmentative and alternative communication (AAC) offers an important set of supports for realizing this critical human right. Yet despite a strong evidence base, many AAC systems are not tailored to the support needs of individuals with visual and/or motor impairments. In this paper we explore the potential for artificial intelligence to interpret body-based communicative movements (including gestures) into speech output. The team, which includes individuals with motor impairments, developed and tested a prototype application that shifts the burden of access off the AAC user and onto the application, through the use of body-based sensors. The effort demonstrated proof of concept that the prototype could learn meaningful idiosyncratic gestures and ignore unintentional ones. Movement sensors interpreted the gestures regardless of the user's or communication device's position, reducing constraints on its usability, potentially increasing the flexibility of AAC access. Dedicated research is needed to confirm this possibility.

ARTICLE HISTORY

Received 8 July 2024
Revised 28 October 2024
Accepted 9 December 2024

KEYWORDS

Unaided AAC; alternative access; gestures; movement; body-based sensors

As noted in the National Joint Committee's Communication Bill of Rights (Brady et al., 2016, p.1), "All people with a disability of any extent or severity have a basic right to affect, through communication, the conditions of their existence." Access to speech-language therapies that promote optimal communication outcomes is also noted to be a fundamental right by the United Nation's Article 19 of the Convention on the Rights of Persons with Disabilities (Article 19, 2016). Yet many individuals with physical or intellectual disabilities have language limitations that prevent them from using speech as their primary mode of communication. For these individuals, assistive communication technologies (augmentative and alternative communication; AAC) offer an important set of supports for realizing this critical human right.

Although AAC is widely-used and evidence-based (Beukelman & Light, 2020, pp. 11–23, 325, 667), there are particular challenges in designing AAC for individuals with visual and/or concomitant motor impairments. Unlike spoken language, in much of aided AAC the vocabulary items are visual (letters, words, symbols) and only a limited number of items can be displayed at a time, since they must be

presented on an external device (such as a tablet or a dedicated AAC device). To maximize available vocabulary, clinicians often place many symbols onto the small display. Although this strategy can be useful for some people, it is a substantial problem for individuals with visual impairments who cannot either see (ocular) or process (cortical) the visual information (Wilkinson et al., 2023). In addition, access to these vocabulary items often involves use of a finger or eye gaze to select a symbol or a limb to activate a switch. These types of repetitive selections may be difficult and fatiguing for individuals with motor disabilities (Sowers & Wilkinson, 2023). As a consequence, traditional methods of accessing AAC that are relatively simple for other individuals are selectively more difficult for those with visual impairment and motor disabilities. There is an urgent need to develop AAC technologies that reduce the visual and motoric burden for such individuals.

In this paper, we propose substantially increasing the flexibility of aided AAC access in part through a reconsideration of the traditional distinction made between aided (i.e., low or high- technology assisted) and unaided (i.e., body-based)

CONTACT Krista M. Wilkinson  kmw22@psu.edu  Department of Communication Sciences & Disorders, the Pennsylvania State University 308 Ford Building, University Park, PA 16802, USA.

This paper was edited by Rose A. Sevcik

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

communication modes. As described in Beukelman and Light (2020), aided communication modes offer the power of symbolic communication that is readily understood by many communication partners, even those who are unfamiliar with the individual using AAC. However, aided modes can be quite limiting in terms of the vocabulary available, speed of message preparation, environmental constraints, and ability to support natural conversations. Unaided communication modes, on the other hand, can involve a diverse range of natural movements that are well within the skill set of the user, and can be rapidly produced with low physical effort. The drawback of unaided modes is that they are often difficult for unfamiliar partners to understand, limiting the range of potential communication partners and necessitating the proximity of a communication partner to the AAC user.

We propose that given contemporary technology, it is both theoretically and practically possible to substantially increase access to aided AAC by leveraging the ability of technology to sense and interpret natural, unaided communicative movements¹ – which we consider to include both conventional gestures like waves or a thumbs up sign, as well as idiosyncratic movements that the user and familiar partners both understand. Virtually all individuals, including those with disabilities, use such communicative movements during daily interactions with familiar partners. On our team, community advisor Emma Elko has cortical (brain-based) visual impairment and uses high-tech aided AAC mounted on her power wheelchair. Emma has a specific unaided “I want” movement that her family understands, but that is hard for unfamiliar partners to interpret. Emma also has a very different unaided movement to indicate excitement that family and close friends understand, but that can be worrisome to unfamiliar partners who may interpret it as indicative of negative emotion or overstimulation. Both of these natural movement patterns carry unique meaning to Emma’s familiar partners, and both are ones that her AAC technology could be trained to read and translate into speech output. Doing so would essentially create a “movement-to-speech” translation akin to existing text-to-speech applications that would expand the range of partners who could understand those specific body-based movements.

By no means are we arguing that this movement-to-speech translation application would replace traditional aided forms of AAC. Rather, the movement-to-speech function would both supplement existing aided AAC and also offer unique additional power for situations in which the traditional aided AAC might not work optimally (that we describe later in this paper). Harnessing unaided inputs as a supplemental means for access to AAC technology will marry the power of the aided symbolic communication with the ease, speed, and unique movements employed by individual users. In so doing, it will shift the burden of access from the user (at least in part) onto the aided AAC technologies themselves. Building flexible technologies that are tailored to the motor and visual skills of individuals with disabilities is well within the capabilities of modern devices and is an active area of research in Human-Computer Interaction (HCI) and accessible computing (Ehtesham-Ul-Haque & Billah, 2023;

Harrison et al., 2011; Kane et al., 2008; Ruiz et al., 2011; Zhang et al., 2020).

Limitations of traditional aided AAC access methods for individuals with visual and/or motor disabilities

Individuals with motor disabilities and/or visual impairment may have difficulty accessing traditionally-presented, visual aided AAC displays either because of movement restrictions or because of difficulties processing visual information. In particular, one challenge of aided AAC relates to the fact that aided AAC largely relies on the presentation of visual symbols on a physically limited space, that is, on the assistive technology screen. Sizable vocabulary offerings are accomplished by large grids filled with symbols, maximizing the number of symbols available on each display. This type of display, however, risks introducing what is called “visual crowding” (van den Berg et al., 2007). Visual crowding occurs when a visual stimulus that might be distinguishable to a viewer when presented alone becomes non-distinguishable when placed close to neighboring symbols (called “flankers” in the vision sciences literature: Whitney & Levi, 2011). Visual crowding can impact access for individuals with either ocular or brain-based visual impairments.

The presence of a motor disability compounds challenges of access. This can in part be due to the reliance of many current AAC systems on access methods that either require precise control of a digit (a finger for selection) or laborious and/or repetitive movements of other body parts (a limb, the head, the eyes) for activation. As noted, precise finger selection is not possible for all persons, and communication via switch control is exceptionally slow and is typically fatiguing due to the need for repeated motor movements (Cook & Polgar, 2015; Wagner & Jackson, 2006; see also Dittmann, 2014). Other direct selection methods include eye or head tracking, but often these alternative methods present constraints due to their design. For example, by design typical alternative access systems have positioning and orientation constraints which require visual attention shift (i.e., to generate a message using eye tracking, the individual must shift their gaze away from the conversational partner to system) to create a message, and may not work well in bright or low light situations (e.g., Elsahar et al., 2019) or outdoor lighting situations (e.g., due to the constraints of eye tracking technology; Binaee et al., 2021).

When challenges with a person’s AAC system are present, individuals will often naturally defer to body-based communicative movements that familiar partners can interpret (Blackstone et al., 2021). As mentioned earlier, this necessitates the communication partner be familiar with the meaning behind the movements and also be in close enough proximity to see those movements. Developing technologies that allow someone to use their natural movements to gain attention, express wants or needs, or promote social closeness has the potential to foster engagement and interaction both with familiar and with unfamiliar communication partners.

Our proposed approach differs from traditional direct selection or switch access in two important ways. First, it responds to diverse movements that are already part of the individual's unaided communication methods. Body-based communicative movements often occur naturally by the individual, and are not explicitly taught to them. Individuals use these movements in real time to convey meaning when with familiar communication partners. Our prototype artificial intelligence algorithms will allow this wide range of existing movement patterns to be interpreted for their intended message, to aid unfamiliar partners in interpreting them. Second, each message is tied uniquely to a single movement or gesture that itself conveys specific meaning. Enabling AAC to interpret movement patterns could allow for expanded use of the voice output of the AAC system in varying positions (e.g., side lying) and environments (e.g., outdoors) that current methods do not readily provide, and for individuals to easily alert someone that their AAC system is not set up properly (e.g., the calibration is off).

Artificial intelligence for interpreting natural communicative movements

Current technologies have begun to incorporate artificial intelligence. For instance, there is some natural gesture recognition capability built into the Android platform via the Project Activate Application (Research at Google, 2021). This application allows the built-in camera to use a person's natural facial gestures (e.g., smile, raised eyebrows, wink) to perform a number of tasks, including phrase generation such as "Hi, how are you?" Furthermore, these technologies can identify movements as specific as a single finger lift or as broad as a left-to-right motion. Uniquely, they can also learn idiosyncratic movements and map user-preferred commands to these movements. When the system adjusts to the individual's intentional movement patterns, rather than the individual having to conform to predetermined actions on screen, the potential for error is reduced, along with the demands placed on the individual (Higginbotham et al., 2007).

However, the utility and user experience of these artificial intelligence algorithms remain unexplored. Additionally, there are gaps in the understanding of how these algorithms are developed, how they can be adapted for AAC users with diverse disabilities, and how they can be seamlessly integrated into existing AAC solutions.

This paper aims to begin to address these knowledge gaps. We present an end-to-end prototype that was informed by three community advisors with motor disabilities and/or CVI over a period of more than eight months. Our prototype enables AAC users to communicate via natural communicative movements. Users wear a small motion sensor, similar to a smartwatch, to make their preferred movements, which our prototype records. Next, it allows users to inspect their recorded movement pattern and interactively create artificial intelligence algorithms with ease. Furthermore, it enables users to fine-tune the algorithm and deploy it into a prototype AAC system running on their smartphone. Importantly, the prototype was deliberately developed to "learn" the

individual, unique communicative movements produced by each communicator, rather than relying on patterns banked from some larger group of people. A communicative movement that might produce the message "I want" for one user might be a very different pattern for that same message produced by a different user. The power of the algorithm is that it can learn the unique movement patterns of each individual.

Co-creating an adaptive prototype for movement-to-speech translation

We divided the co-creation of our adaptive AAC prototype into four phases. First, we created a team with domain experts and target users and chose a starting point. Second, we identified challenges in recognizing unaided input for AAC users based on feedback from the first trial by one of our community advisors. Third, we addressed these challenges through human-centered artificial intelligence practices. Finally, we reflected on the development process. Each phase built upon the previous one, allowing us to develop and refine our prototype iteratively while maintaining a focus on the needs and experiences of AAC users.

Creating a team representing the three pillars of evidence-based practice

Wilkinson et al. (2023) synthesized existing research on visual characteristics in individuals with brain-based (cortical) visual impairment (CVI). That paper incorporated the three pillars of evidence-based practice, in which suggestions about AAC system design were informed not just by scientists and clinicians, but by the lived experiences of two of the authors, that is, a young woman who has CVI herself as well as her parent/advocate. The aim of that paper was to provide guidelines for how these perspectives might specifically guide design of traditional aided AAC approaches in which direct selection is the method for access.

The current effort shifts gears to focus on supplementing traditional forms of aided AAC, through development and initial field testing of artificial intelligence algorithms that allow AAC technology to interpret diverse natural movements that are already part of individuals' communication repertoire. The team in the current effort again represents the three pillars of evidence-based practice, that is scientific evidence, clinical expertise, and the lived experiences of individuals with motor impairments. The team for the current paper reflects external and internal expertise across a wide range of scientific, clinical, and first-person experiences regarding critical aspects of the project. Authors Billah and Namdeu bring expertise in programming of smart technologies that are accessible for individuals with low vision (Billah et al., 2018; Momotaz & Billah, 2021). Authors Wilkinson and Sowers bring scholarly and clinical expertise in designing AAC that is responsive to the unique visual processing skills of users with developmental and/or intellectual disabilities (Sowers & Wilkinson, 2023, 2024; Wilkinson et al., 2023; in press). Author Redmon brings extensive experience with applications of assistive technology within educational settings. Authors Elko, Williams, and

Case bring first- or second-person lived experiences regarding the supports that promote successful outcomes for individuals with diverse motor disabilities; these authors represent a range of profiles of skills and will henceforth be described as “community advisors.” As will be described, each team member/coauthor has contributed from the initial conceptualization of the priorities of algorithm development to the methods for accomplishing those goals.

Choosing a starting point, on the basis of community advisor recommendations

At the outset, the team discussed two possible avenues for testing the potential of artificial intelligence for detecting natural communicative movements; first, the use of the built-in camera on virtually all hardware (both dedicated devices or general-purpose ones like iPads) to detect movements, and second, the use of a sensor placed somewhere on the body that would send movement information via Bluetooth to the sensor detector, using a dedicated smartphone.

The community advisors were unanimously excited about the second avenue, specifically, the use of the sensor, as the starting point. Their enthusiasm related to the enhanced power a sensor would have not just to detect the movement itself (which a camera could also do) but to do so at a distance from the smartphone that receives the input and generates the intended spoken message. This is important because it allows the communication to occur without the voice output technology being directly next to the AAC user, a point we return to later.

Each of our community advisors contributed their perspectives on the potential for this application within their own lives. Coauthor Williams is a wheelchair user who communicates via traditional aided AAC using direct selection. He states this about the potential role of incorporating the proposed type of technology within his active lifestyle:

I have been a long-time AAC user starting from manual systems on paper to the tablet based systems I am speaking to you today on. Each system takes a lot of engagement from me and other users, to where we are speaking, that is not only the primary focus of what we are doing, but the methods we use limit us to a single task at a time, which is using my communication device. I am an active person who likes to do activities that are adventurous like water skiing, swimming, and taking care of my infant daughter. Like many people I need to be able to communicate some quick needs during tasks, where my device is either not feasible to use or is not safe to use (I have yet to see an AAC device that is not affected by coming into contact with water). Anyway, I joined this project to expand my communication toolset to include a way to speak without having to focus on a screen, and do it in places where I can literally communicate with a flip of my wrist.

It is important to note that our proposals are not (yet) that the sensor could be used in the midst of an activity like waterskiing. We do not know if the sensor would be able to transmit reliable movement information while an individual is actually on the waterski in the wake of the traveling speedboat. However, that limitation is just as true of users of

natural speech as well – when a person who uses speech is skiing in the wake of the speedboat, even shouted speech will likely not be adequate to be heard by listeners on the speedboat. Rather, our argument is that the sensor would allow communication during pauses in any activity, to request help with something (adjusting the line) or signaling readiness to start the boat moving. This point is reinforced by coauthor Elko, who comments on how the sensors could augment the low-tech partner-assisted scanning that is the current mode of communication during breaks in snow skiing for Emma:

Emma is a snow skier. I usually have to bring paper pages of her talker and do partner assisted scanning when we take breaks from skiing. But with the system, she would be able to communicate on the fly...

Coauthor Case brings the perspective of an individual who has used a wheelchair his whole life. Though Case uses natural speech for communication, he represents the perspective of an advocate who has experienced the challenges of accessing support services for individuals with physical disabilities. In an early meeting, Case pointed out one of the potential applications of the sensor on a practical level: When the AAC technology can interpret the movement patterns, the need for aide-by-aide onboarding of personal care aides in understanding the meaning of an individual's unique and specific gestures or movements is essentially removed, as the aide can learn from the voice output of the technology. As he notes:

For those of us with motor impairments who have a lot of personal care aide staff turnover... the sensor technology is able to make the onboarding process, in ... training a new staff member, much easier and smoother because [it] will enable us to communicate what our needs and our routines are in such a way that for somebody who is new, they will become more familiar with what our movements mean.... This is particularly salient in terms of how this artificial intelligence-based technology can be employed because nationally, the statistics say that the staff turnover rate for personal care aides in the US is 77–100 percent. (Friedman, 2021; Rollinson et al., 2023).

Identifying challenges and solutions in recognizing unaided input

In the published AAC literature, there are several conceptual or research development discussions of the potential of various forms of gesture recognition to support AAC users with multiple disabilities (e.g., Boster & McCarthy, 2017; Fager, et al. 2019; Higginbotham, et al, 2007; Sennott, et al., 2019). It appears that while the concept of using natural movement patterns to supplement AAC systems has been considered, the current developments in technology suggest that now is the time to move forward in putting them to use in AAC systems.

Outside of AAC, smartphones offer some unaided input recognition as accessibility features. However, the usefulness of these techniques and their potential integration into existing AAC solutions remain largely unexplored (Boster & McCarthy, 2017). Our analysis of existing techniques and

unaided input recognition algorithms revealed four key challenges in developing personalized artificial intelligence for AAC. In this context, artificial intelligence refers to real-time gesture recognition algorithms that can interpret and translate a user's natural communicative movements into spoken outputs. We outline the challenges here and then describe how we addressed them in the following section.

The first challenge is designing personalized artificial intelligence with limited data. The algorithm must adapt to user-specific movements using only a few training samples, unlike generic movement recognition systems trained on large datasets (Wobbrock et al., 2007). It must also differentiate between intentional and unintentional movements, effectively "delimiting" the start and finish of intentional communicative movements. Additionally, the algorithm should be flexible, capable of fusing data from multiple sensors when available. For instance, users might wear sensors on their wrist or upper arms of their dominant hand, non-dominant hand, or both. They could place sensors on nearby surfaces like a wheelchair headrest. This flexibility allows the system to adapt to user needs and preferences while maintaining efficient performance.

The second challenge is developing a low-compute algorithm for real-time operation. To be effective, the system must recognize communicative movements as quickly as they are made. This requires the algorithm to run directly on the AAC device, such as a smartphone or iPad, without consuming excessive battery power. This process is described later in this section.

The third challenge is creating user-friendly interfaces to promote users' agency in developing the algorithm. The interface must be easy to use and designed for non-technical users, enabling them to record their preferred communicative movements (i.e., the template movements), mark the beginning and ending of the recorded templates, and recognize

that the sensor data is representing the templates movements in a meaningful way. The interface should also allow users to create, test, and iteratively improve the algorithm's performance until its movement detection accuracy reaches an acceptable threshold.

The fourth challenge is integrating artificial intelligence into AAC technology. The algorithm must seamlessly integrate with users' AAC devices, allowing them to create multiple movement patterns and contexts, and map each movement to an intent based on their situation. For example, a wave in a public space might translate to "Hi, how are you?" while the same movement in a private setting could mean "No, I don't want it."

Designing human-centered personalized movement pattern recognizer and user interfaces

To address the aforementioned challenges, we broke down the process of communicating through unaided hand movements into four phases: (1) Initial Setup; (2) Recording Template Gestures/Movements; (3) Gesture/Movement Visualization, Annotation, and Model Creation; and 4) Real Time Gesture/Movement Detection and Model Deployment. Figure 1 illustrates these phases with a step-by-step breakdown, showing how we distributed the workload across the sensor device, smartphones, and PC. Some steps are exclusively performed by users (Steps 1, 3.a, 10.a), while others can be completed by either users or their aides. Throughout each phase, we consulted with our community advisors and refined our approach based on their feedback.

Initial setup: the IMU sensor

Figure 2 illustrates the setup. The AAC user (shown in Panel C) wears a generic Inertial Measurement Unit (IMU) sensor.

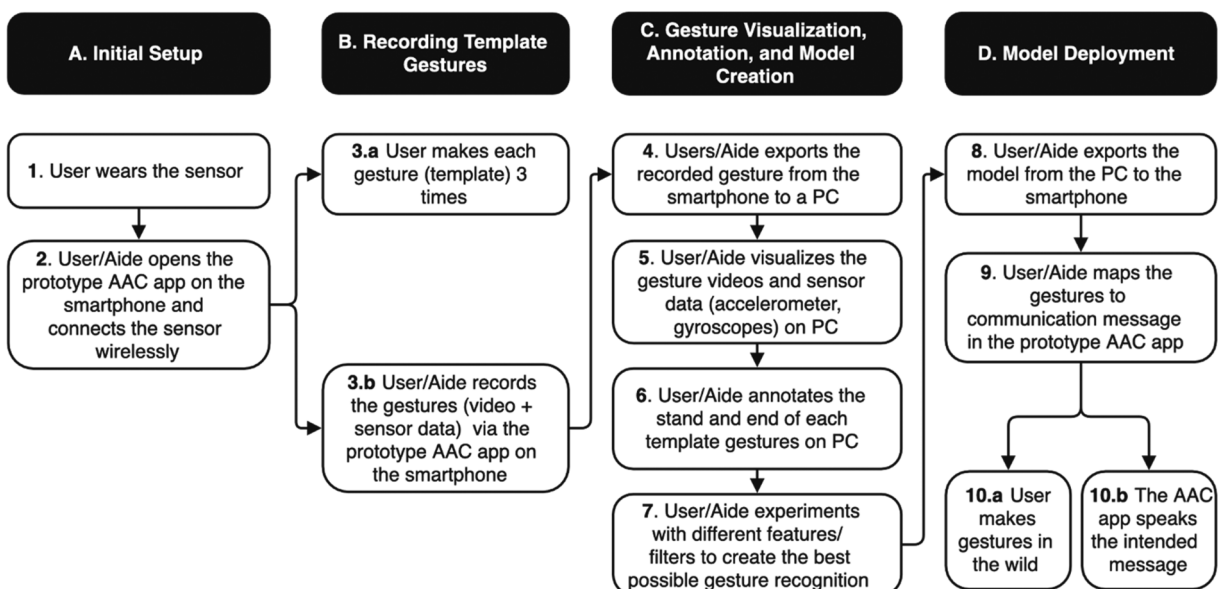


Figure 1. A flow diagram illustrating the step-by-step process from sensor setup to the final use of our prototype AAC application for unaided communication. Note: The term "gesture" in this figure refers to any communication movement, including both conventional gestures (thumbs up, wave) and the idiosyncratic communication movements that are understood to have specific meaning by an individual communicator and their familiar partners.

We used a Movesense² sensor (shown in Panel A), though any comparable IMU sensor could be used. It provides 9-axis movement: accelerometer readings along X, Y, and Z axes; gyroscope readings along the 3 axes; and magnetometer readings along 3 axes. We only used accelerometer and gyroscope readings, ignoring magnetometer readings due to noise. The sensor is water and shock-proof and connects to the smartphone via Bluetooth Low Energy (BLE). It weighs 9.4g with battery, has a diameter of 36.6mm, and a body thickness of 8mm. Panel D shows the Bluetooth screen to connect to one or more sensors. It displays a list of previously connected devices for quick reconnection. Users can also scan for new devices, ensuring the app can connect to various sensors as needed.

Recording template gestures/movements

In this phase, the user wears the sensor and makes the movement patterns that they want the algorithm to recognize and interpret. Panel B of Figure 2 shows an aide, in this case a family member of the AAC user (who can be a layperson), recording the user's preferred movements using our app on the smartphone. Depending on their abilities, users can record themselves by tapping the record button on the app and placing the smartphone on a stationary surface or holding it with their other hand.

Panel E of Figure 2 shows the Data Recording screen. This interface resembles a camera, enabling simultaneous recording of video and sensor data. This dual recording capability is essential for creating a comprehensive dataset. The video

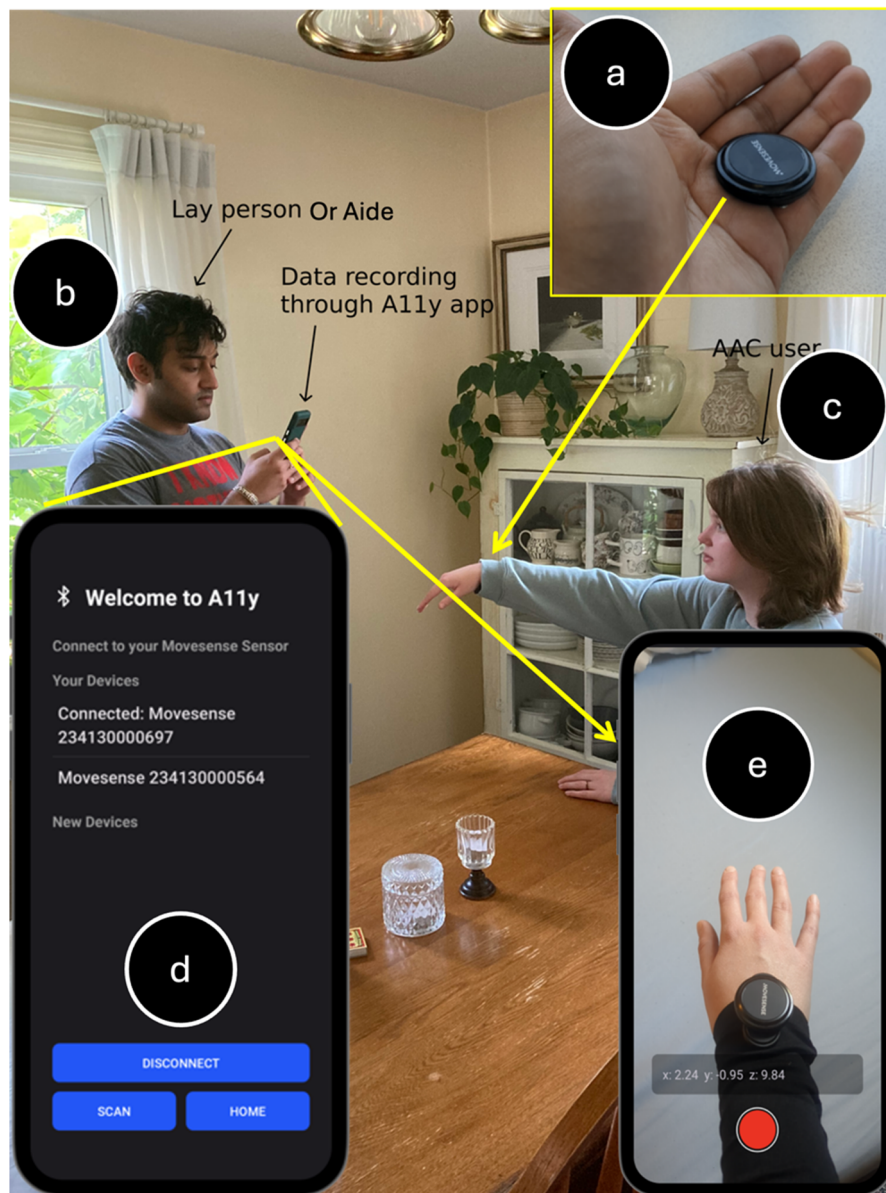


Figure 2. Illustration of sensor placement and operation. *Note:* Panel A: The Movesense sensor used in our system; Panel B: An aide (in this case, a layperson) recording the movement/gesture data of an AAC user wearing the sensor on her wrist; Panel C: The AAC user; Panel D: Bluetooth screen on our prototype AAC app for connecting to one or multiple sensors; Panel E: Recording screen on the app for collecting sensor data (accelerometer and gyroscope) as well as the associated video.

Table 1. A set of template communicative movements used by one co-designer who uses AAC

Template communicative movements (intent)	
Circular hand motion (contextual)	Fist open, then up and down (contextual)
Covering the ears with both hands (dislike)	Thumb between the fingers and waving (washroom)
One finger up and down (contextual)	Waving hands while showing the palm (greeting)
Rubbing a hand on the chest (contextual)	Both hands in the air (happy)
Cross hands up and down (contextual)	Extending hand outward (contextual)
Extending hand and calling (want)	Hands over the head (contextual)
Push away (disagreement)	Hands over the ears (contextual)
Hands over the lap (idle)	Resting hand (idle)

data serves as an important visual reference for later gesture/movement annotation. During recording sessions, the application captures sensor data at a rate of 52Hz. This frequency is suitable for recognizing human activities; it ensures that the captured movement patterns have fine granularity while being efficient enough to avoid causing app lag or excessive battery consumption.

Users begin by recording *template* gestures/movements, which include at least one “idle” movement to indicate their default state that the algorithm should ignore. For unaided communication, they must also record one or more “non-idle” communicative movement patterns they intend to use (as shown in Table 1). To account for natural variability in execution, users perform each template movement three times, with each repetition lasting approximately two seconds. This approach helps the algorithm capture slight differences that may occur when making the same movement at different times. Recording these templates can be accomplished in a variety of ways. One of our advisors prepared a list of movements they use and performed each three times in one session. For another advisor, their aide, who is a familiar communication partner, identified natural communicative movements that were often used for intentional communication. Those movements were then elicited via natural means as they occurred. For example, the three template recordings were made over a snack time where the individual readily uses their movement pattern repeatedly to indicate that they wanted another bite of food. The ability to trim the recordings (see section below) allow for the recording to be active to “catch” the natural occurrence of the movement pattern as it happens.

We limited the number of trials to three, striking a balance between gathering sufficient data and minimizing user fatigue. This user-centered approach stands in contrast to traditional deep learning techniques, which often require millions of data points for training. Indeed, our data requirement is insignificant by comparison. We prioritize users’ abilities and comfort, and collect just enough information for effective gesture or communicative movement recognition.

Gesture/movement visualization, annotation, and model creation

The specifics of the visualizer, annotator, and model builder are described in detail below. As mentioned above (the third challenge), it was critical to create a user interface that could

be readily manipulated and accessible for users or aides at any level of technical background, meaning that it would be easy to understand and use. As will be described later, field testing with our advisors is indicating that the interface is readily understood and usable in real-world conditions.

The workflow begins after the recording of the movement pattern on a smartphone is completed. Users or their aide then export this raw data to a PC, where they utilize our integrated gesture/movement visualizer and annotator, illustrated in Figure 3, along with the interactive model builder, illustrated in Figure 4. Note that we used a PC because it has larger display and higher computing resources. It is possible to perform all operations on the smartphone app with the sacrifice of some features.

The Integrated Gesture/Movement Visualizer interface (Figure 3) displays video recordings (Panel A) alongside corresponding accelerometer and gyroscope readings in three axes: X in red, Y in green, and Z in blue (Panel B). This integrated display enables users to correlate hand movements with sensor readings, recognize patterns, and pinpoint the start and end of movement patterns. For instance, the accelerometer readings along the Z-axis in panel B are substantially lower (-10.0 m/s^2 or below) compared to X or Y axes. This is because the Z-axis reading represents the gravitational force, which is close to -9.8 m/s^2 when the sensor is at rest. The negative value indicates that gravity acts in the opposite direction of the positive Z-axis. Moreover, X and Z readings show more variation than Y, as the rotation, in this example (a circular movement shown in panel A), primarily occurs around the Y-axis. The Y-axis, in this case, is aligned with the long axis of the user’s forearm. This alignment means that when the user rotates their wrist, the sensor experiences more change in the X and Z directions than in the Y direction. This graphical display of the sensor data in relation to the video of the physical movements enhance users’ understanding of how their communicative movements translate into data patterns. Consequently, it increases their confidence in interpreting the data and in the process of algorithm creation.

The Gesture/Movement Annotation interface allows users to segment the video and mark beginnings and endings of movement patterns, illustrated in Panel D of Figure 3. Users annotate different trials of template movements, saving them with identical prefixes (template name) but unique suffixes (trial ID). For instance, Panel C of Figure 3 shows three circular movement template trials annotated as *circle_1*, *circle_2*, and *circle_3*. This annotation process enables precise definition of each movement’s temporal boundaries, crucial for accurate model training. Annotating multiple trials of the same movement pattern accounts for natural variations in performance, enhancing the recognition model’s robustness. The interface also supports editing of previously annotated movements, so that users can refine annotations after initial reviews or correct mistakes to improve the quality of the templates. Thus, our interactive process can enhance users’ understanding of movement variability and increase their engagement in data preparation and their confidence in creating the gesture/movement recognition model.

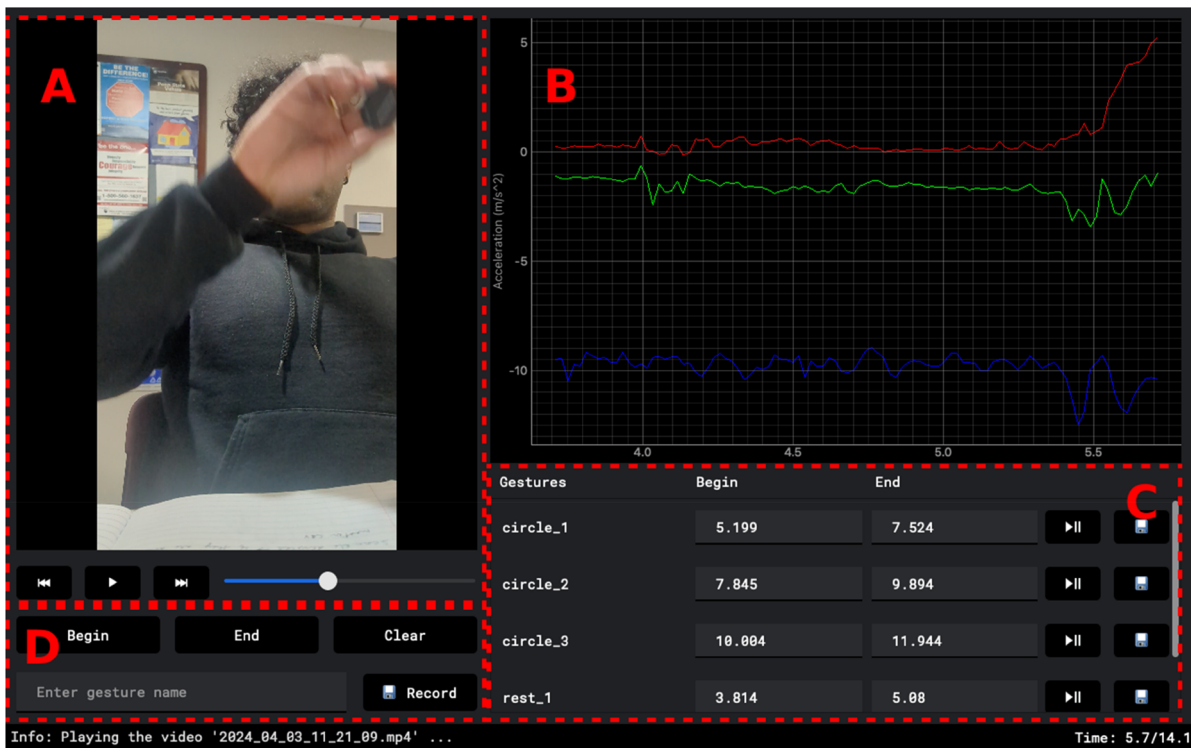


Figure 3. Illustration of the integrated communication movement visualizer. Note: Panel A: Play/pause control for gesture video; Panel B: timeline of sensor data (accelerometer or gyroscope in 3 axes) representing a gesture; Panel C: movement/gesture annotation with beginning and ending timestamps; Panel D: save annotations function.

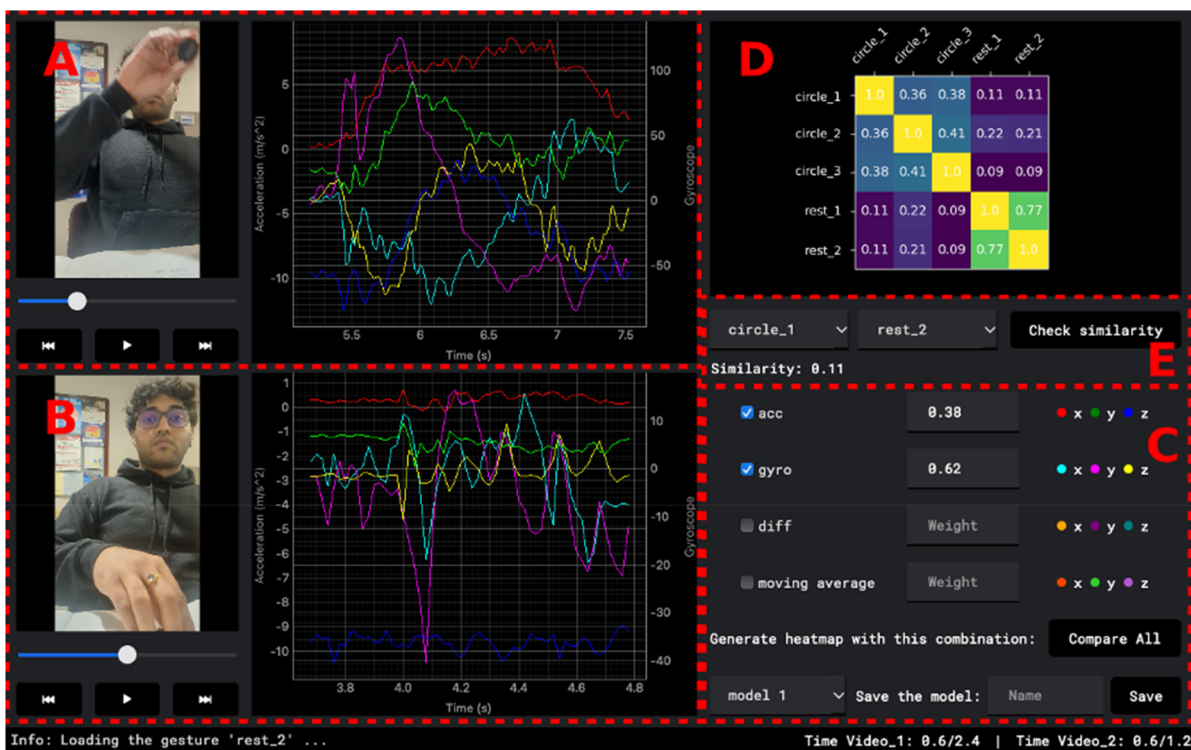


Figure 4. Illustration of interactive model builder. Note: Panel A: template gesture/movement; Panel B: candidate gesture/movement; Panel C: adjustable attributes, filters, and weights for model building; Panel D: model performance (i.e., the confusion matrix) shown as a heatmap; Panel E: the similarity score between two gestures/movement loaded in A and B.

After completing the annotation, users employ an interactive model builder (Figure 4), to build and refine the gesture/movement recognition model. This interface allows users to load

gesture pairs for comparison: typically, an instance of a template movement (Panel A) and an instance of another template gesture, henceforth called the *candidate* movement (Panel B).

Users start by visually comparing the video and the line graphs of their sensor readings (Panels A and B). Next, they construct the model by applying various features and filters (Panel C). These include raw sensor data (accelerometer or “acc” and gyroscope or “gyro”), a difference filter or “diff” (calculating the difference between current and immediate past readings), and a moving average filter (computing the average of a feature over N timestamps). We chose these computationally inexpensive yet powerful filters to capture both instantaneous variations in gestural patterns and long-term trends.

As users apply these features and filters, they can visualize the resulting transformations (Panel B) and similarity scores (Panel E). Similarity scores, ranging from 0 to 1 (1 being most similar), are calculated using the longest common subsequence (LCSS) metric. The primary goal is to select or unselect features and filters, and adjust their weights, to maximize similarity within the same template movements while minimizing similarity between different templates.

The model’s performance is displayed as a color-coded confusion matrix (Panel D) for easy interpretation. Each row and column represent a template movement, with cell (i, j) indicating the similarity score between template movements i and j . High similarity appears as bright yellow, low similarity as purple blue. Diagonal cells are brightest, representing identical movement patterns. Rows or columns with many bright cells indicate potential misrecognition. Users can iteratively adjust the model parameters to reduce these misrecognitions, or in some cases, remove overly similar templates. Once satisfied, users save the best-performing model.

This interactive process helps users understand the challenges in recognition of communicative movement patterns. For example, one of our advisors attempted to produce movements that were distinguished only by differences in thumb placements (index-middle fingers versus middle-ring fingers). These subtle movements proved difficult for the algorithm to differentiate using only accelerometer and gyroscope data. Through such interactions, this advisor learned to avoid movements with such subtle distinctions, increasing their confidence in model building and demonstrating the interface’s empowering nature.

Finally, a scoring function can be used to measure the similarity between any two movement patterns. We applied the Longest Common Subsequence (LCSS) algorithm, a dynamic programming technique widely used in various domains including DNA sequencing, text correction, and gesture recognition (Nguyen-Dinh et al., 2012). LCSS identifies the longest sequence of matching elements between two sequences, allowing for gaps while maintaining order. This algorithm is suitable for our model as it can handle variations in movement speed and small interruptions between slightly different trials of the same template movement.

The key challenge in applying LCSS is converting continuous time series sensor readings (e.g., acc, gyro) into a sequence of discrete symbols. To achieve this, we employed a k-Means algorithm with the number of clusters, k , equal to 20 (empirically determined). This algorithm assigns the sensor readings (e.g., acc readings along three axes) at a given time to one of 20 clusters, labeled C_1 to C_{20} . Figure 5 illustrates

this process. Panel A illustrates the 2-second sliding window (red rectangle) that increments by 1 second (dotted rectangle). This results in approximately 50% overlap between consecutive windows, sufficient for maintaining temporal continuity across windows. Panel B provides an illustration of the k-Means algorithm converting continuous time series sensor data into a discrete symbol sequence. Each symbol denotes a cluster ID (C_1 to C_{20}). The X-axis represents time, while the Y-axis shows sensor readings along three axes (color-coded). Note that sensor readings at 0.16s and 0.22s are slightly different but fall within the same cluster, C_7 . Ultimately, sensor readings are converted to a sequence of cluster IDs: $C_3C_4C_1C_2 \dots C_{12}C_7$. We precomputed the cluster sequences for all template movements following this process. These sequences are then used during real-time gesture/movement recognition, described next.

Real time gesture/movement detection and model deployment

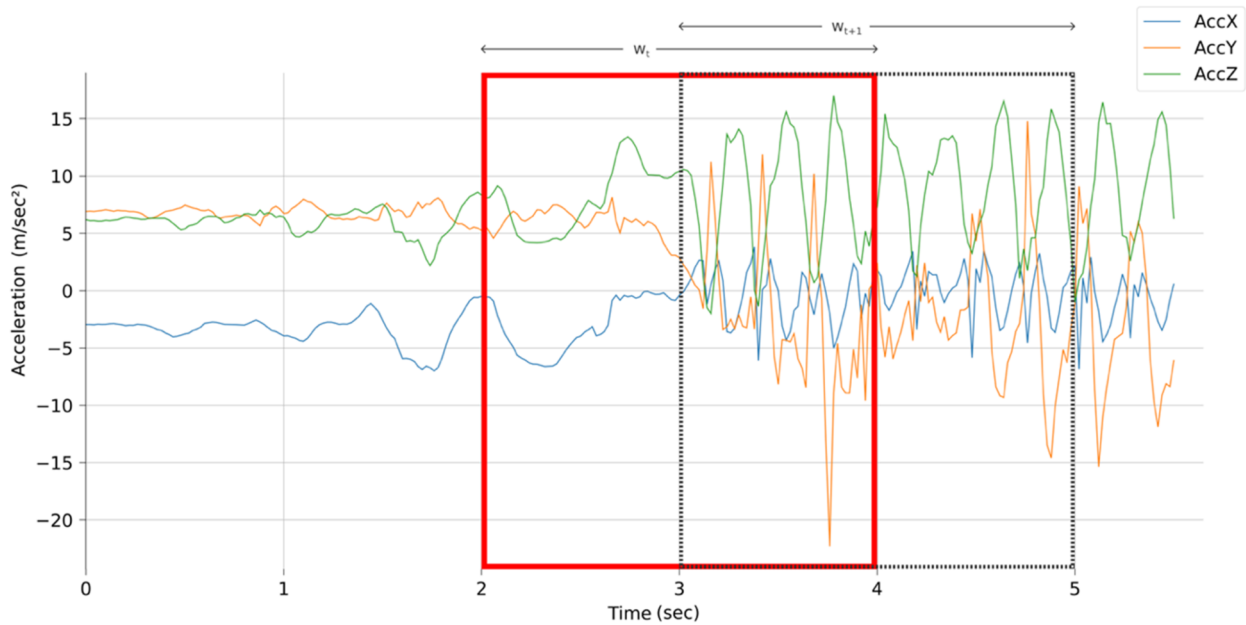
Once the model is exported to a smartphone and the AAC screen is active, the application receives a continuous data stream from the sensor in real-time. This is illustrated in Panel A of Figure 6. The model runs in the background to recognize any template movements. This process is challenging because users can make a movement at any time. To reduce computational cost, we invoke the gesture/movement recognition algorithm at 1 second intervals on a sliding window of length 2 seconds, as was illustrated in Panel A of Figure 5.

We convert the sensor data within the current sliding window into a sequence of cluster IDs using the k-Means algorithm with 20 precomputed clusters. We then compute the LCSS scores between this sequence and all precomputed template sequences. This comparison generates an array of similarity scores, indicating how closely the incoming sequence matches each stored template. To recognize a movement, we select the highest score from the array and compare it against a pre-set threshold (0.4, empirically determined). If the highest score surpasses this threshold, the current movement is considered successfully recognized.

Future direction: template movements across contexts

A key lesson is allowing users to map the same template movement to different communication messages depending on their current context. For example, users can map the “waving hands” communicative movement to “Hi” in public settings but “No” in private settings. Figure 6 shows our current interface. Users can edit the message associated with each movement pattern (Panel B). These messages are automatically inserted into the AAC app’s text box and read out by a Text-to-Speech (TTS) engine when a registered movement pattern is recognized (Panel C). Users can select a model from the dropdown menu labeled “Model.” This feature enables users to choose from different models designed for various contexts, rather than the system automatically determining the user’s current context. Although this functionality is not yet implemented, the intended mechanism is as follows: when a user selects a different model, the system will update both the registered communicative movement

Panel A:



Panel B:

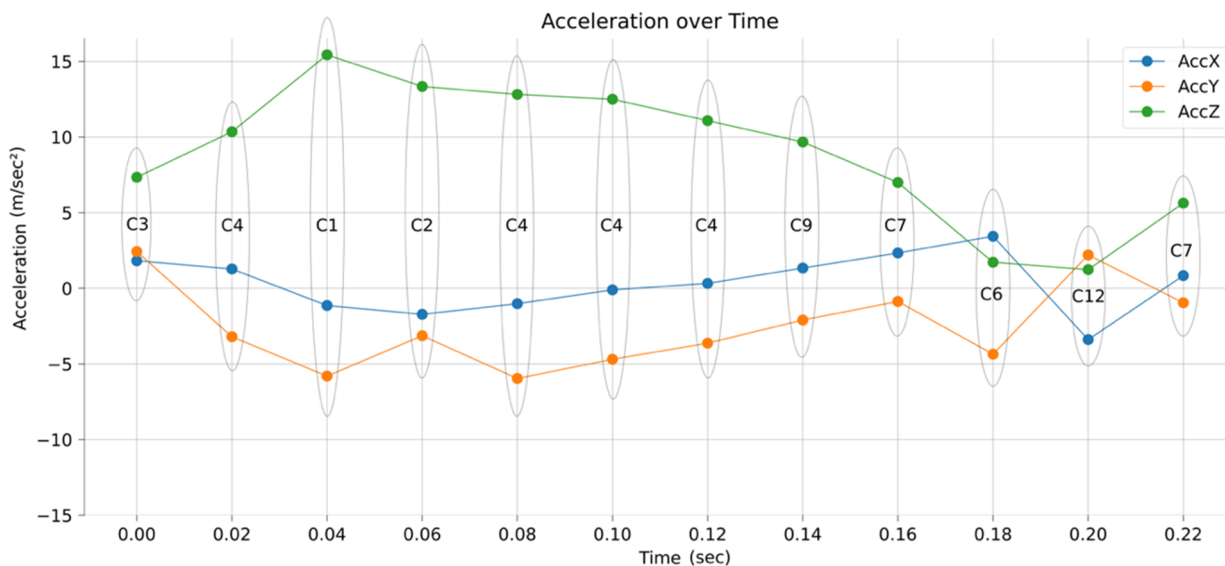


Figure 5. Illustration of sliding window (Panel A) and conversion of continuous sensor data to discrete cluster ID sequence (Panel B).

patterns and their associated predefined text for readout. This implementation is incremental, and we have already developed the foundational components necessary to achieve this feature.

Computational efficiency of the gesture/movement recognition system

As noted earlier, a challenge we addressed was allowing the system to operate rapidly on a smartphone without consuming excessive battery power. The gesture/movement recognition system is computationally efficient by design. We use k-Means clustering ($k=20$) to reduce 6-axis IMU data to integers, minimizing memory use and simplifying computations.

It needs only 90seconds of training data for 15 movements, far less than deep learning methods. Our LCSS algorithm performs 10,816 integer comparisons per 2-second window, versus 64,896 floating-point operations for similar algorithms like Dynamic Time Warping (DTW). With a 10KB memory footprint, it is so far leaner than traditional neural networks. This enables real-time processing of 104 data points per recognition, allowing responsive, near realtime performance.

Preliminary proof of concept evaluation and summary

Our advisors assisted with the development by trialing our prototype, which encompassed the sensor, smartphone app, visualizer and model creation tools, and the gesture/movement

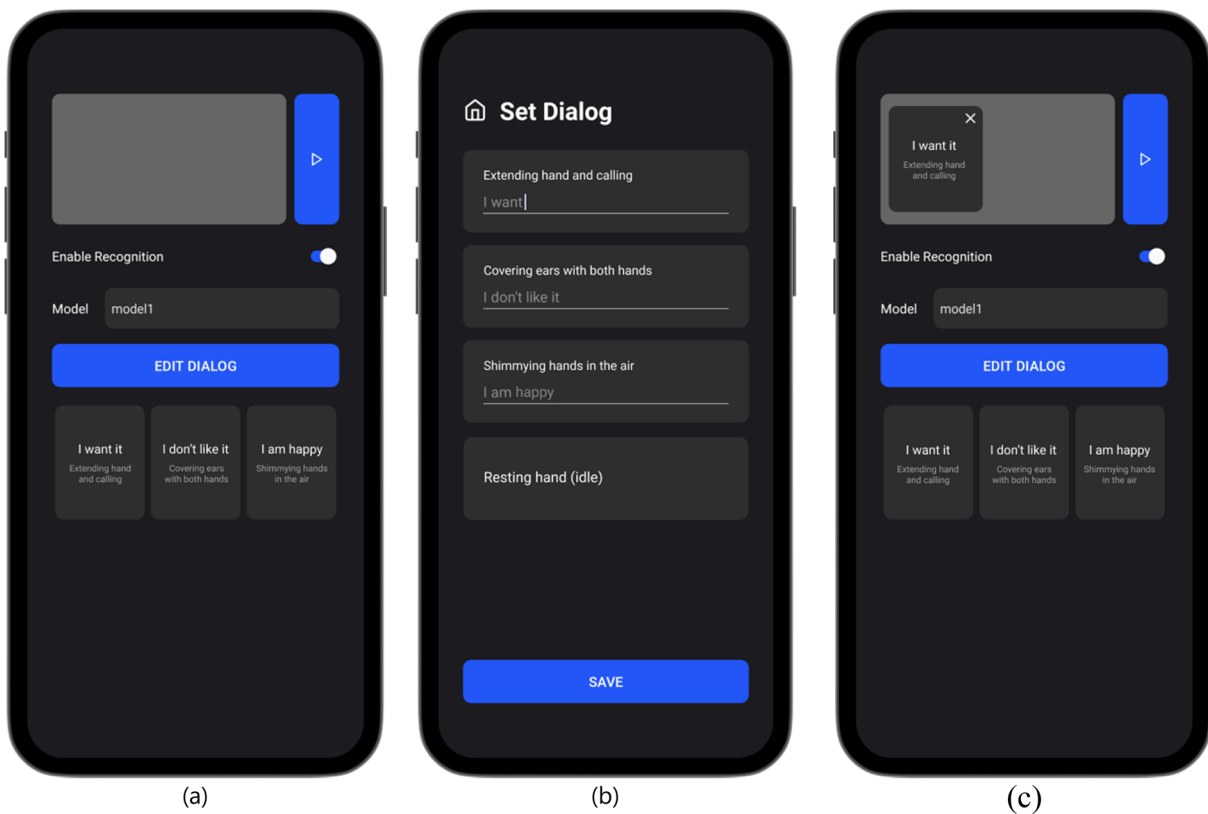


Figure 6. Illustration of unaided movement/gesture integration into existing AAC applications. *Note:* Panel A: A prototype AAC app that receives sensor data continuously and runs the movement/gesture detection algorithm in the background for a potential match. Panel B: The app allows users to edit communication messages associated with communication movements. Panel C: The message (e.g., “I want it”) is automatically selected in the AAC textbox when the user makes the associated movement (e.g., extending one hand and calling movement) and the app correctly detects it.

recognition model/algorithm. Our advisors used the prototype at each different phase of its maturity, offering insightful feedback that guided our iterative development process.

Changes made over the development process

Initially, we only recorded sensor data for each communicative movement, but quickly realized after coauthor Case tested it for us that interpreting this data without visual reference was impractical. This led us to incorporate video recording using an external camera. However, this approach introduced new challenges when tested by community advisor Emma: lack of synchronization between the camera’s frame rate and the sensor’s sampling rate, and difficulties in annotation due to manual video manipulation required to align with sensor data.

We addressed these issues by improving the movement recording process. Our app now simultaneously records both video and sensor data with perfect synchrony. The integrated gesture/movement visualizer and annotator resolves the challenge of manually aligning movement pattern data with video frames to identify the start and end of movements. This improvement also eliminated the need for sensor calibration for each user - a process that involved asking users to make canonical movements along X, Y, or Z axes, which proved challenging for both of our advisors who tried it. This preliminary evaluation has been crucial in refining our prototype but also highlights its potential to meet the practical needs of AAC users. It also highlights the importance of

user-centered design in developing assistive technologies, where real-world testing and user feedback drive continuous improvement.

Preliminary field testing of the prototype

The current prototype is being used by our advisor Emma in home and field-based communication situations. [Figure 7](#) illustrates this real-world application, showing a community advisor expressing “I want” during a physical therapy appointment within which she was not in close proximity to her AAC system due to the nature of the activity. The sensor is mounted on her left hand, with speech output produced by a phone (not visible in the image). Wearing the sensor, she was able to communicate to her physical therapist that she wanted an item that was out of her reach. Her physical therapist was able to respond to the voice output that interpreted the body-based communication movement and respond to the request.

Another example occurred during an outing to the AAC user’s local pickleball courts. As coauthor Elko describes it:

I play pickleball, and Emma loves to spectate...Typically, I try to keep a line of sight to Emma so that if she needs something, I can recognize her need, because we don’t take her volume all the way up on her talker because it’s inconsiderate to other players. So this day, I was able to move around the courts on typical play. I had the cell phone with me, and Emma had her body-based sensors, and she was able to communicate with me without my having to be able to see her, and I was able to be far away from her on the court.

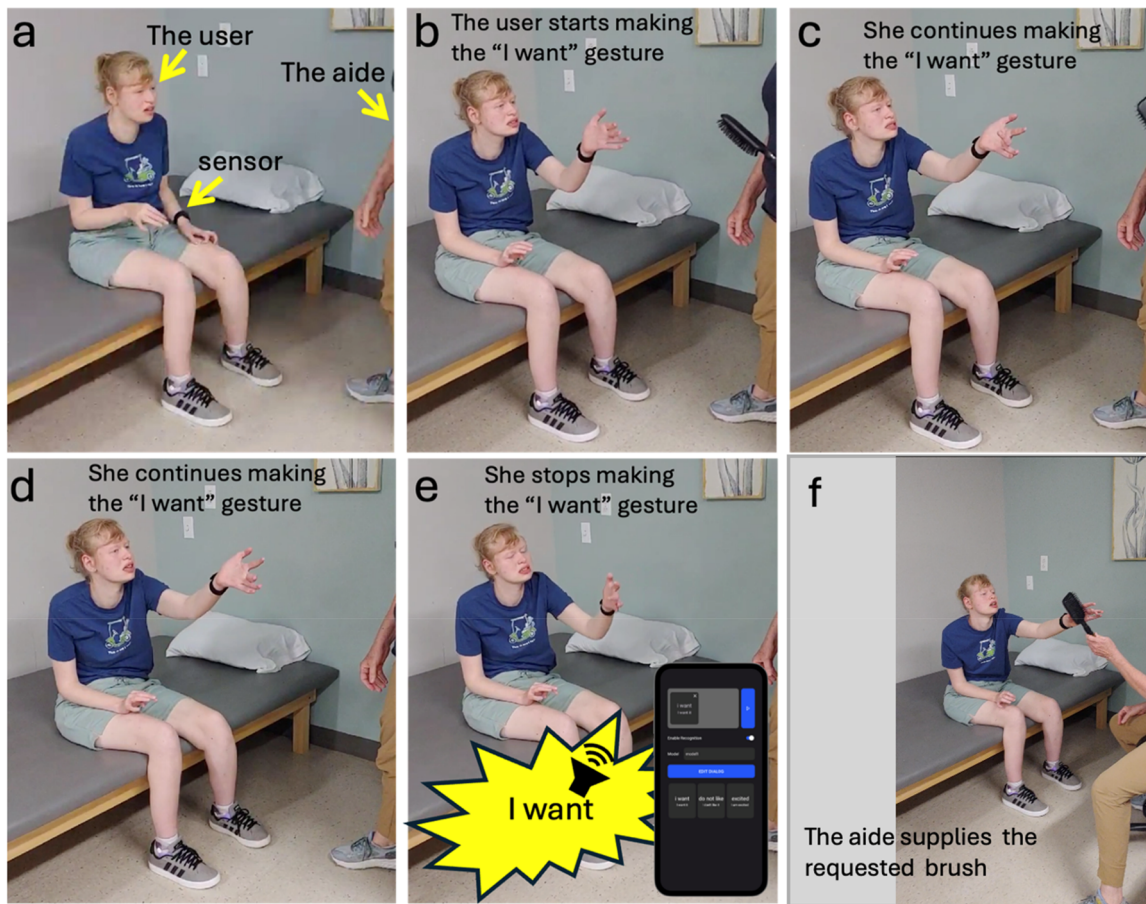


Figure 7. An illustration of AAC user with cortical visual impairment communicating with her physical therapist through her idiosyncratic “want” communication movement. Note: The user is sitting on a bench wearing the sensor on her left wrist.

Importantly, at all times Emma still had access to her traditional aided AAC, with which she could communicate with partners who were proximal to her. As Emma is a regular on the pickleball courts, everybody knows her and her forms of aided AAC-based communication. Thus, Emma still had her traditional AAC to communicate with the partners who remain proximal to her, but now had an added level of independence from her mother – which is appropriate for a young woman of her age (early 20s).

Summary

We have discussed development of artificial intelligence algorithms that are capable of interpreting individual-specific unaided gestures or idiosyncratic communicative movements. We offered proof of concept of the operation of demonstration algorithms developed by our team, including our community advisors (and coauthors) with motor disabilities who were an integral part of system development from conceptualization through implementation. Our system is intended to be human-centered, use-inspired, and readily-programmed, to empower both the user and their communication partners who may be involved in AAC services. It is designed specifically to capitalize on movements that are already in each individual’s repertoire and that already communicate specific messages. The sensor is worn by the individual communicator, so message production is independently authored by that individual and

thus is not facilitated in any way by the partner. The very next step is to conduct research on how well the prototype operates, with individuals with a wide range of motor and visual impairments, to ensure that the algorithms can operate across a wide range of motor abilities and for individuals with CVI.

Notes

¹The team had a lively discussion over whether to use the term “gesture” or “movement pattern.” The liveliness of that discussion reflected the likelihood that the term “gesture” is interpreted differently in the fields of speech-language-pathology (and associated fields) and human-computer interaction. In the main text, we have sought to define that the patterns we are focusing on here include both conventional gestures, such as waves or thumbs up, but also the idiosyncratic movement patterns that many individuals who use AAC employ with familiar partners. For the remainder of this paper, we have settled on “communicative movements” and “movement patterns” as our preferred terms, but also use “gesture/movement” in cases where the text is referring specifically to the labels common in human-computer interaction.

²<https://www.movesense.com/product/movesense-medical-mdr/>

Acknowledgements

The ideas laid out in this paper were developed in part through two grant awards to authors Wilkinson, Sowers, and Billah from the Penn State Center for Biodevices and a “Big Ideas” grant from Penn State Center for Socially Responsible Artificial Intelligence, on which authors Elko, Case, and Williams serve as community advisors. Authorship order reflects relative contribution for the first set of authors, and that the work is being developed by the final three authors. The first author is co-funded through the Penn State Social

Science Research Institute. We are indebted to Emma Elko for providing the insights about AAC strategies that served as the initial foundation for this paper. The fifth author was supported, in part, by funding from the Penn State AAC Leadership Project, a doctoral training grant funded by the U.S. Department of Education, grant H325D220021.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Center for Biodevices, Pennsylvania State University; Penn State Center for Socially Responsible Artificial Intelligence.

References

- Article 19 (2016). *Background paper: Freedom of expression and information and persons with disabilities (policy brief)*. United Nations Article 19 of the convention on the rights of persons with disabilities. <https://www.article19.org/resources.php/resource/38511/en/consultation-open-principles-on-freedom-of-expression-and-information-and-persons-with-disabilities>
- Beukelman, D. R., Light, J. C. (2020). Augmentative & alternative communication: Supporting children and adults with complex communication needs. Brookes Publishing. <http://ebookcentral.proquest.com/lib/pensu/detail.action?docID=6229697>
- Billah, S. M., Ashok, V., Porter, D. E., & Ramakrishnan, I. V. (2018). SteeringWheel: A locality-preserving magnification interface for low vision web browsing. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI'18) (pp. 1–13). Association for Computing Machinery. <https://doi.org/10.1145/3173574.3173594>
- Binaee, K., Sinnott, C., Capurro, K. J., MacNeilage, P., & Lescroart, M. D. (2021). Pupil tracking under direct sunlight. ETRA '21 Adjunct; ACM Symposium on Eye Tracking Research and Applications (vol. 18, pp. 1–4).
- Blackstone, S., Luo, F., Canchola, J., Wilkinson, K., & Roman-Lantzy, C. (2021). Children with cortical visual impairment and complex communication needs: Identifying gaps between needs and current clinical practice. *Language, Speech, and Hearing Services in Schools*, 52(2), 612–629. https://doi.org/10.1044/2020_LSHSS-20-00088
- Boster, J. B., & McCarthy, J. W. (2017). When you can't touch a touch screen. *Seminars in Speech and Language*, 38, 286–296.
- Brady, N. C., Bruce, S., Goldman, A., Erickson, K., Mineo, B., Ogletree, B. T., Paul, D., Ronski, M., Sevcik, R., Siegel, E., Schoonover, J., Snell, M., Sylvester, L., & Wilkinson, K. (2016). Communication services and supports for individuals with severe disabilities: Guidance for assessment and intervention. *American Journal on Intellectual and Developmental Disabilities*, 121(2), 121–138. <https://doi.org/10.1352/1944-7558-121.2.121>
- Cook, A., & Polgar, J. (2015). *Assistive technologies: Principles and practice* (4th ed.). Mosby.
- Dittmann, A. T. (2014). The role of body movement in communication. In *Nonverbal behavior and communication* (pp. 37–64). Psychology Press.
- Ehtesham-Ul-Haque, M., & Billah, S. M. (2023). Abacus gestures: A large set of math-based usable finger-counting gestures for mid-air interactions. *Proceedings of the ACM on Interactive Mobile Wearable and Ubiquitous Technologies*. Wearable Ubiquitous Technology.
- Elsahar, Y., Hu, S., Bouazza-Marouf, K., Kerr, D., & Mansor, A. (2019). Augmentative and alternative communication (AAC) advances: A review of configurations for individuals with a speech disability. *Sensors*, 2219(8), 1911. <https://doi.org/10.3390/s19081911>
- Fager, S., Fried-Oken, M., Jakobs, T., & Beukelman, D. R. (2019). New and emerging access technologies for adults with complex communication needs and severe motor impairments: State of the science. *Augmentative and Alternative Communication*, 35(1), 13–25. <https://doi.org/10.1080/07434618.2018.1556730>
- Friedman, C. (2021). The impact of direct support professional turnover on the health and safety of people with intellectual and developmental disabilities. *Inclusion*, 9(1), 63–73. <https://doi.org/10.1352/2326-6988-9.1.63>
- Harrison, C., Benko, H., Wilson, A. D. (2011). OmniTouch: Wearable multi-touch interaction everywhere. In *Proceedings of the 24th annual ACM symposium on user interface software and technology* (pp. 441–450). <https://doi.org/10.1145/2047196.2047255>
- Higginbotham, D. J., Shane, H., Russell, S., & Caves, K. (2007). Access to AAC: Present, past, and future. *Augmentative and Alternative Communication*, 23(3), 243–257. <https://doi.org/10.1080/07434610701571058>
- Kane, S. K., Bigham, J. P., Wobbrock, J. O. (2008). Slide rule: Making mobile touch screens accessible to blind people using multi-touch interaction techniques. *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility* (pp. 73–80).
- Momotaz, F., & Billah, S. M. (2021). Tilt-explore: Making tilt gestures usable for low-vision smartphone users. In The 34th Annual ACM Symposium on User Interface Software and Technology (UIST'21) (pp. 1154–1168). Association for Computing Machinery.
- Nguyen-Dinh, L. V., Roggen, D., Calatroni, A., Tröster, G. (2012). Improving online gesture recognition with template matching methods in accelerometer data. In *2012 12th International Conference on Intelligent Systems Design and Applications (ISDA)* (pp. 831–836). IEEE.
- Research at Google (2021). *Project activate*. Google Play Store. <https://play.google.com/store/apps/details?id=com.google.android.apps.vision.switches>
- Rollinson, J., Bandini, J., Feistal, K., Gittens, A., Key, M., González, I., Kong, W., Ruder, T., & Etchegaray J. M. (2023). An evaluation of a multisite, health systems-based direct care worker retention program: Key findings and recommendations. *Rand Health Quarterly*, 10(2), 4.
- Ruiz, J., Li, Y., Lank, E. (2011). User-defined motion gestures for mobile interaction. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 197–206).
- Sennott, S. C., Akagi, L., Lee, M., & Rhodes, A. (2019). AAC and artificial intelligence (AI). *Topics in Language Disorders*, 39(4), 389–403. <https://doi.org/10.1097/TLD.0000000000000197>
- Sowers, D. J., & Wilkinson, K. M. (2023). Demands associated with an AAC system in relation to alternative forms access for individuals with motor impairments. *American Journal of Speech-Language Pathology*, 32(1), 37–54. https://doi.org/10.1044/2022_AJSLP-22-00006
- Sowers, D. J., & Wilkinson, K. M. (2024). SLPs' Approach to alternative access assessment for AAC for children with motor impairments. *American Journal of Speech-Language Pathology*, 33(2), 774–790. https://doi.org/10.1044/2023_AJSLP-23-00129
- van den Berg, R., Roerdink, J., & Cornelissen, F. W. (2007). On the generality of crowding: Visual crowding in size, saturation, and hue compared to orientation. *Journal of Vision*, 7(2), 14–11. <https://doi.org/10.1167/7.2.14>
- Wagner, B. T., & Jackson, H. M. (2006). Developmental memory capacity resources of typical children retrieving Picture Communication Symbols using direct selection and visual linear scanning with fixed communication displays. *Journal of Speech, Language, and Hearing Research*, 49(1), 113–126. [https://doi.org/10.1044/1092-4388\(2006\)009](https://doi.org/10.1044/1092-4388(2006)009)
- Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, 15(4), 160–168. <https://doi.org/10.1016/j.tics.2011.02.005>
- Wilkinson, K. M., Elko, L., Elko, E., McCarty, T., Sowers, D., Blackstone, S., & Roman-Lantzy, C. (2023). An evidence-based team approach to the design of AAC interventions for individuals with Cortical Visual Impairment. *American Journal of Speech-Language Pathology*, 32(5), 1939–1960. https://doi.org/10.1044/2023_AJSLP-22-00397
- Wilkinson, K. M., Sainburg, R., & Sowers, D. (in press). A pilot study of the influence of visual-perceptual features of augmentative and alternative communication displays on motor behavior by individuals with Down syndrome. *Augmentative and Alternative Communication*.
- Wobbrock, J. O., Wilson, A. D., Li, Y. (2007). Gestures without libraries, toolkits or training: A \$1 recognizer for user interface prototypes. *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology* (pp. 159–168).
- Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C. L., & Grundmann, M. (2020). Mediapipe hands: On-device real-time hand tracking. *arXiv preprint arXiv:2006.10214*.