Swype AI: A Multimodal Voice and Gesture Control System for Accessible Education

Dhanvinkumar Ganeshkumar, Thomas Jefferson High School for Science and Technology

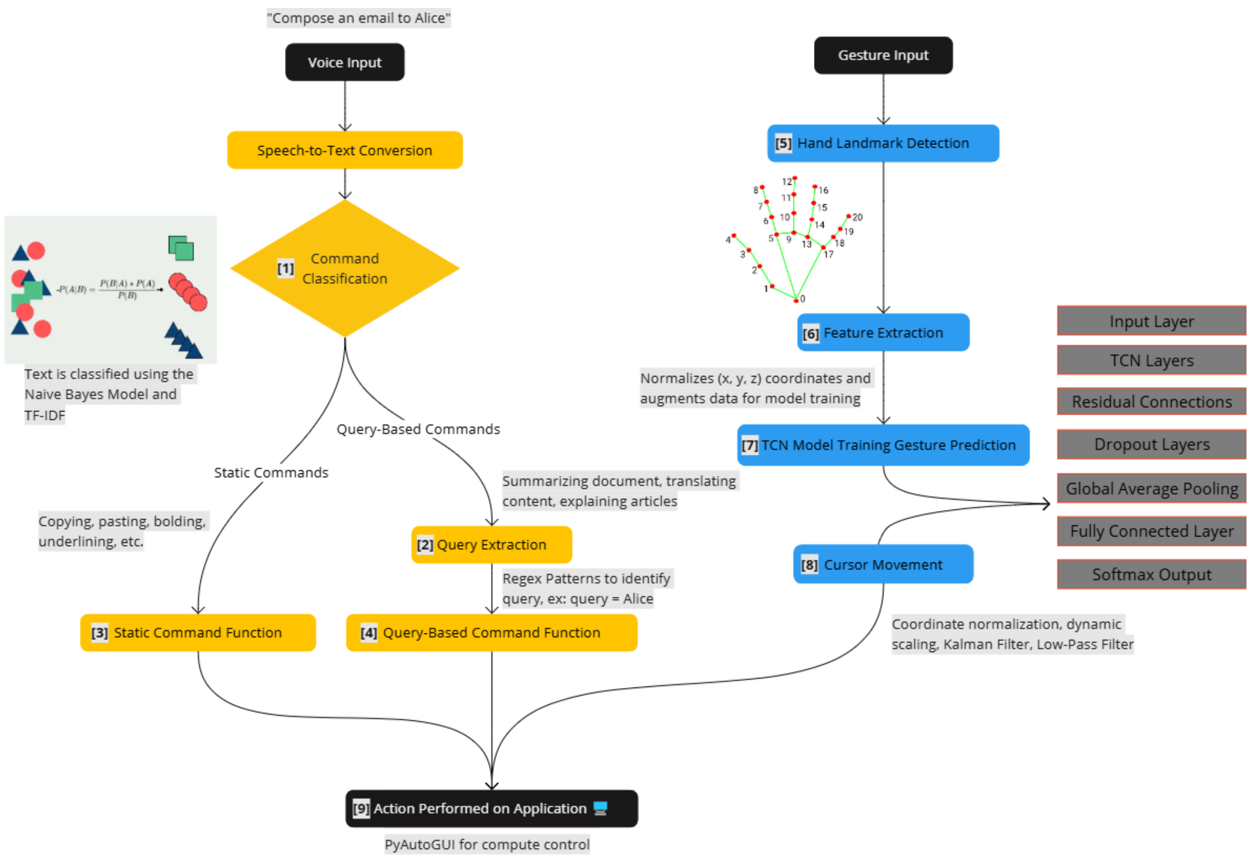
Globally, more than 1.3 billion people, approximately 16% of the world's population, live with some form of disability. Many, including those with conditions like Parkinson's, ALS, or muscular dystrophy, face significant barriers to computer use and digital access (World Health Organization, 2022). Traditional input methods like keyboards and mice often become inaccessible, creating barriers to digital learning and independent task completion. These challenges are further compounded by perceptual and cognitive impairments, including deficits in smooth pursuit eye movements, which hinder object tracking, and impaired predictive motor control, which affects hand-eye coordination necessary for typing or controlling a cursor (Chakravarthi et al., 2023). Without adaptive technologies, many individuals risk digital exclusion, missing essential opportunities for learning, communication, and participation. Digital exclusion translates directly into lost instructional hours, reduced classroom engagement, and inequities in test-taking conditions.

As education systems increasingly rely on digital platforms, equitable computer access becomes a prerequisite for success. Online assignments, virtual classrooms, learning management systems (e.g., Google Classroom, Canvas), and digital testing platforms are standard across K–12 and higher education. Students with motor impairments face disproportionate disadvantages when they cannot easily navigate these systems. Thus, addressing technological accessibility is essential to closing the education gap and ensuring that all students can engage meaningfully in modern learning environments. This paper, submitted under Track 2: Innovative Applications of AI, introduces Swype AI, a multimodal hands-free computer interface designed to support students with motor impairments in educational environments.

While touchless computing via voice or gesture shows promise, current solutions face key limitations. Speech recognition for users with dysarthria yields only 50–60% accuracy (Masina et al., 2020), and gesture systems like Kinect remain costly, hardware-dependent, and error-prone under varying lighting (Ermolina & Tiberius, 2021).

System Overview and Architecture

Swype AI addresses these gaps through a real-time software system that combines natural voice and gesture control to replace traditional peripherals. It runs on consumer laptops without requiring specialized hardware, and the architecture is structured around two core pipelines:



The voice recognition system (left) uses Term Frequency–Inverse Document Frequency (TF-IDF) vectorization to extract important keywords from spoken commands, transforming them into numerical feature vectors. A Naïve Bayes classifier then predicts user intent based on these features. To handle dynamic inputs, such as names or search queries, the system applies regular expressions (Regex) for flexible parameter extraction. Commands are categorized into 36 static actions (e.g., "bold text," "select all") and 10 query-based actions (e.g., "compose email to Alice," "search cats"), trained on a custom dataset of 1,530 labeled examples. Preprocessing steps include lowercasing, special character removal, and tokenization to optimize recognition accuracy.

The gesture control system (right) leverages MediaPipe Hands to detect 21 key landmarks per frame on the user's hand, generating a 63-feature vector (x, y, z coordinates). These sequential frames are classified using a Temporal Convolutional Network (TCN), capable of recognizing both static gestures (such as a closed fist) and dynamic gestures (such as a wave upward). Kalman Filters are applied to the predicted hand trajectories, and the model was trained on 10 sequences of 30 frames per gesture.

Swype AI was designed for lightweight, real-time use on consumer-grade hardware. Voice samples (n = 1,530) were recorded using a standard laptop microphone in quiet indoor environments. The dataset includes inputs from five speakers across diverse accent backgrounds, including two non-native English speakers. Gesture sequences (n = 300) were captured via webcam under variable lighting to simulate real-world use. A Naïve Bayes classifier was selected due to its efficiency and low memory overhead, ideal for deployment on typical Intel Core i5 laptops with 8GB RAM. Swype maintained <30 ms latency with <350MB peak RAM usage during inference.

Preliminary Results:

Module	Accuracy
Voice Command Intent	93.1%
Static and Dynamic Gestures	88.3%

Prototype testing showed 93.1% voice intent accuracy and 88.3% gesture recognition accuracy, and demonstrations included document editing in Google Docs, composing and sending emails through natural voice commands, and conducting web searches with integrated gesture navigation. Current limitations include reduced recognition accuracy for users with severe dysarthria and misclassification of dynamic gestures in low light. Future work includes developing a fine-tuned, lightweight transformer model to improve robustness, expanding the dataset to better represent accented and disfluent speech, and exploring lightweight onboarding calibration to adapt models to individual speech and gesture patterns.

Educational Impact and Use Cases

Swype AI enables students with motor disabilities to independently complete digital learning tasks, including writing essays, conducting online research, and navigating LMS platforms like Google Classroom and Canvas. Additional features support classroom tasks, including gesture shortcuts for annotating PDFs or highlighting passages in e-books, voice macros for quickly accessing assignments or grades, and compatibility with exam software that restricts external hardware. In virtual classrooms, gestures such as raising a hand can trigger participation tools in Zoom or Google Meet, ensuring full engagement in both in-person and online settings.

A 15-student usability study is planned in collaboration with assistive organizations. The study cohort will include participants with dysarthria and varying motor impairments to evaluate performance in the intended user population. Each participant will complete three representative tasks (e.g., locating an assignment, opening an online quiz, submitting answers) using Swype. The study will measure task-completion time and administer the System Usability Scale (SUS) to assess perceived usability and learning effectiveness (Lewis & Sauro, 2017). We hypothesize that Swype will improve task efficiency by at least 30% while maintaining a SUS score above the standard usability threshold of 70. All inference runs locally on-device; no audio or video data is transmitted externally to maintain user privacy.

Preliminary outreach included conversations with over 15 accessibility organizations, including the Parkinson's Foundation, the National Multiple Sclerosis Society, and representatives from the U.S. Department of Health and Human Services. Feedback emphasized the need for flexible multimodal systems that could adapt to users' changing motor and speech abilities, reinforcing Swype's design choices. To promote transparency and accessibility, Swype AI will be released under the MIT License as an open-source project following the Summit.

References

Chakravarthi, B., M., P., Imandi, R., & B N, P. K. (2023). *A comprehensive review of Leap Motion Controller-based hand gesture datasets* (pp. 1–7). https://doi.org/10.1109/NEleX59773.2023.10421030

Data Basecamp. (2022). *What is the Naive Bayes algorithm?* https://databasecamp.de/en/ml/naive-bayes-algorithm

Ermolina, A., & Tiberius, V. (2021). Voice-controlled intelligent personal assistants in health care: International Delphi study. *Journal of Medical Internet Research, 23*(4), e25312. https://doi.org/10.2196/25312

Lewis, J. R., & Sauro, J. (2017). Revisiting the factor structure of the System Usability Scale. *Journal of Usability Studies*, 12(4), 183–192.

Marais, M., Brown, D. L., Connan, J., & Boby, A. Z. (2022, August). *An evaluation of hand-based algorithms for sign language recognition*. In *2022 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)* (pp. 1–7). IEEE. https://doi.org/10.1109/icABCD54961.2022.9856310

Masina, F., Orso, V., Pluchino, P., Dainese, G., Volpato, S., Nelini, C., Mapelli, D., Spagnolli, A., & Gamberini, L. (2020). Investigating the accessibility of voice assistants with impaired users: Mixed methods study. *Journal of Medical Internet Research, 22*(9), e18431. https://doi.org/10.2196/18431

Murad, B., & Alasadi, A. (2024). Advancements and challenges in hand gesture recognition: A comprehensive review. *Iraqi Journal for Electrical and Electronic Engineering, 20*(2), 154–164. https://doi.org/10.37917/ijeee.20.2.13

Rish, I. (2001). An empirical study of the Naïve Bayes classifier. In *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence* (Vol. 3, pp. 41–46).

Robertson, S. (2004). Understanding inverse document frequency: On theoretical arguments for IDF. *Journal of Documentation, 60*(5), 503–520. https://doi.org/10.1108/00220410410560582

World Health Organization. (2022). *Disability and health*. https://www.who.int/news-room/fact-sheets/detail/disability-and-health