FingerSpeller: Camera-Free Text Entry Using Smart Rings for American Sign Language Fingerspelling Recognition

David Martin*
dmartin305@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Zikang Leng*
zleng7@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Tan Gemicioglu
tgemici@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Jon Womack
jonwomack@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Jocelyn Heath
jheath38@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Bill Neubauer
wneubauer4@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Hyeokhyen Kwon
hyeokhyen.kwon@dbmi.emory.edu
Emory University
Atlanta, Georgia, USA

Thomas Plöetz
thomas.ploetz@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Thad Starner
thadstarner@gmail.com
Georgia Institute of Technology
Atlanta, Georgia, USA

ABSTRACT

Camera-based text entry using American Sign Language (ASL) fingerspelling has become more feasible due to recent advancements in recognition technology. However, there are numerous situations where camera-based text entry may not be ideal or acceptable. To address this, we present FingerSpeller, a solution that enables camera-free text entry using smart rings. FingerSpeller utilizes accelerometers embedded in five smart rings from TapStrap, a commercially available wearable keyboard, to track finger motion and recognize fingerspelling. A Hidden Markov Model (HMM) based backend with continuous Gaussian modeling facilitates accurate recognition as evaluated in a real-world deployment. In offline isolated word recognition experiments conducted on a 1,164-word dictionary, FingerSpeller achieves an average character accuracy of 91% and word accuracy of 87% across three participants. Furthermore, we demonstrate that the system can be downsized to only two rings while maintaining an accuracy level of approximately 90% compared to the original configuration. This reduction in form factor enhances user comfort and significantly improves the overall usability of the system.

CCS CONCEPTS

• **Human-centered computing → Interaction devices**; **Accessibility technologies**; **Mobile devices**.

KEYWORDS

wearables, fingerspelling, text entry, sign language, smart rings

*__Both authors contributed equally to this work.__

ASSETS '23, October 22–25, 2023, New York, NY, USA

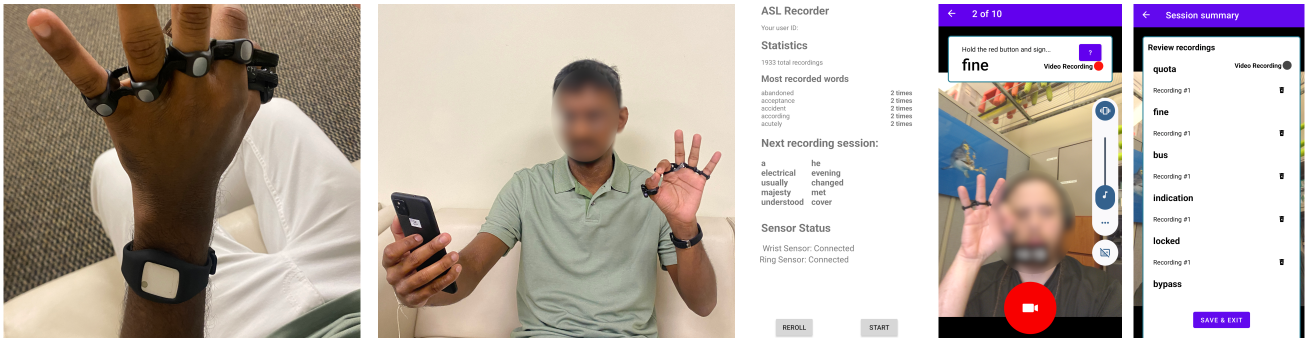
ACM ISBN 979-8-4007-0220-4/23/10.
https://doi.org/10.1145/3597638.3614491

ACM Reference Format:
David Martin, Zikang Leng, Tan Gemicioglu, Jon Womack, Jocelyn Heath, Bill Neubauer, Hyeokhyen Kwon, Thomas Plöetz, and Thad Starner. 2023. FingerSpeller: Camera-Free Text Entry Using Smart Rings for American Sign Language Fingerspelling Recognition. In *The 25th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '23), October 22–25, 2023, New York, NY, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3597638.3614491

1 INTRODUCTION

Automatically recognizing American Sign Language (ASL) for its approx. 500,000 Deaf and Hard-of-Hearing (DHH) signers [23] has been an area of interest for decades [4] but little practical progress has been made with respect to providing access to ASL recognition technology for the DHH community. Serving as an example for this disconnect, gloves equipped with sensors are often meant by researchers to be used to translate sign languages into text or speech – yet they are typically not the first choice for the Deaf community for convenience, practical, and utility reasons [4, 7, 14]. Firstly, gloves that are intended to translate signs into text or speech fail to capture crucial linguistic features, particularly Non-Manual Markers (NMM) [6, 7]. NMM include expressions such as eye gaze, shifting the torso, lip movements, eyebrow movements and head tilts; which–when used–can fundamentally change the meaning of a sentence [2, 6]. Secondly, sign language gloves are supposed to facilitate communication between DHH and hearing individuals. However the burden is placed upon the DHH individual who has to carry around the gloves, batteries, and computer. Wearing such typically bulky equipment restricts the natural movements and speed of signing [7].

In contrast, an often requested feature by the DHH community are interfaces that recognize sign language [12] and connect with personal assistants such as Alexa. However, there are privacy concerns when using camera-based personal assistants, like the personal assistant picking up on signs not meant for the device [11]. In addition, smartphone cameras for sign recognition present additional obstacles such as restrictive field-of-views (which may



(a) (b) (c) (d) (e)

Figure 1: Overview of our data collection process using ring sensors and push-to-sign collection app. (a) Participant wearing the sensors. (b) Participant collecting finger spelling data. (c) Statistics page providing key data collection statistics. (d) Recording page where participants hold down the recording button while fingerspelling the specified word. (e) Summary page enabling participants to view and delete recordings.

not fully capture the signer in the frame), poor lighting in the environment, power consumption of recording and uploading video, and difficulty holding the camera while signing [21].

A potential compromise to aforementioned complications and privacy concerns is the use of smart-rings for the recognition of fingerspelling using the American Manual Alphabet (AMA). Such a system alleviates privacy concerns, bypasses complications with smartphone cameras, provides a more comfortable form factor and can provide general text-entry which has a variety of use-cases. Use-cases can potentially be texting, entering a query into a search engine, filling out text fields, taking notes on the go, learning and/improving fingerspelling, etc.

We propose FingerSpeller which utilize a 5 smart ring system, not as a final system but as a means to test the feasibility of fingerspelling recognition as well as experimenting with omissions of rings to test accuracy with each ring omission. This paves the way towards a smaller, more comfortable, and socially acceptable form factor for its user. Our contributions are four-fold: we introduce FingerSpeller, a camera-free system for fingerspelling recognition; we evaluate FingerSpeller on data collected from three participants and demonstrate that FingerSpeller can achieve an average 91% character accuracy and 87% word accuracy; we further demonstrate that the system can be reduced to 2 rings only while still maintaining approximately 90% relative accuracy, allowing for increased user comfort; and we introduce a dataset of FingerSpeller input that includes 42,288 letters in 6,984 words collected by three participants [1]

2 RELATED WORK

Fingerspelling: Fingerspelling typically is used for names, nouns, and adjectives in approximately 77% of fingerspelt words [25] and can be at rates of 60-96 words per minute (WPM) [28]. In contrast, in a study with 37,000 smartphone typist, it was found the average smartphone typing rate was 36 WPM [26], warranting the exploration of text entry by fingerspelling as a significantly faster alternative. In recent years substantial work has been made towards camera based fingerspelling recognition such as the release of the ChicagoFSWild+ dataset [31] which has motivated the creations of different recognition models and techniques [5, 9, 19, 27, 30]. In addition, on May 10th, 2023- Kaggle launched a competition [1] to develop camera based recognition models for fingerspelling recognition to be used as an alternative keyboard for text entry for DHH individuals. Despite the significant advancements in camera-based fingerspelling recognition, error rates remain surprisingly high for recognition at normal fingerspelling speed, and the exploration of camera-free fingerspelling recognition remains a potentially desirable domain, especially if inconspicuous, off-the-shelf hardware could be made available.

Alternative Text-Entry: While a growing number of interfaces are designed for text entry, few are capable of silent, on-the-go input. The Twiddler [20], a one-handed keyboard, enabled text entry at 60 WPM using chorded entry on a keypad. However, such interfaces encumber one hand with the device, which has to be stowed when text entry is not being performed. As an alternative, interfaces have begun to use subtle wearable form factors such as earables [32], glasses [15] and smartwatches [13] for the text entry system. Silent speech interfaces have been proposed to enable hands-free text entry at the speed of speech, with mobile input emerging as a priority in recent research [17, 32, 33]. However, silent speech is not always applicable and requires sensors to be near the mouth and face.

Sensing with Smart Rings: The development of smart rings dates to before 2000 utilizing embedded accelerometers to detect finger taps [8]. Since then, a multitude of smart ring research has focused on creating smaller form factors [16, 24] as well as creating accurate recognition of gesture [36, 38, 39]. We are now at a point where the hardware and software of smart rings can now be used to explore new use cases. In a position paper by Gheran et al., the scientific community was urged to investigate smart rings as an assistive technology for individuals with motor impairments [10]. This call to action has spurred research into gesture input methods for upper-body motor impairments [35]. ssLOTR [40] explored custom-designed ring sensors for recognizing the AMA. However, the focus of the classification was limited to individual

[1] https://github.com/ZikangLeng/FingerSpeller

letters, leaving word-level recognition and the potential application of ring sensors for text-entry unexplored.

3 FINGERSPELLER

FingerSpeller is designed to identify isolated words fingerspelt using AMA, utilizing accelerometer signals captured by ring sensors. The recorded data is subsequently processed through a backend recognizer to translate the fingerspelled input into text. To personalize the recognizer, we created a push-to-sign collection app that enables users to collect their own fingerspelling data and train a personalized recognizer.

In our system, we utilized TapStrap [34], a commercially available ring sensor system. Each ring of TapStrap is equipped with a tri-axial accelerometer, sampled at a rate of 200 Hz, with the ability to transmit the accelerometer signal to smartphones via Bluetooth. Figure 1(a) shows a participant wearing the sensors. In the AMA, each letter is represented by a distinct handshape. From qualitative observation in early experiments, we found that the unique finger positions and orientations in each handshape can be captured by the gravitation pull applied to accelerometer signals deriving a unique signal profile for each letter, showing that our choice of ring sensors can capture the variations in how different letters are signed.

For personalized data collection, we developed a push-to-sign collection app with three main pages: statistics, recording, and summary. Figure 1(b) shows a user collecting fingerspelling data using our app. The statistics page displays the meta information (Figure 1(c)). During the recording process, users press and hold the recording button on the app's recording page (Figure 1(d)) to start capturing their finger-spelling of the specified words. The button was released after each word is recorded. After recording, users can review the recorded words and corresponding video recordings on the summary page (Figure 1(e)). They can choose to delete any recordings with inaccurate finger spellings to manage data quality. The app saves the timestamps, accelerometer signals, and word start and end times at the end of each session

The collected 15-dimensional time series from five fingers are first segmented according to the start and end timestep of each word. To recognize the word in a given sensor segment, we develop letter-level Hidden Markov Models (HMM) following previous work [17]. Specifically, we further segment word-level sensor time series into letter-level time series. We assume that each letter in a word is fingerspelled in an equal duration of time and later take advantage of Viterbi alignment to better segment the letters.

With the collection of letter-level sensor segments for 26 English letters, we trained an 8-state HMM having a left-to-right topology with no skip transitions. The emissions are modeled using a Gaussian Mixture Model (GMM) with single mixture components, whose parameters were estimated using the Baum-Welch algorithm [3]. After training the 26 letter-level HMMs, we use them to infer words from word-level time-segments using the Viterbi recognizer from the Hidden Markov Model Toolkit (HTK) [37]. This generates the recognized letter sequence, for which we compute the Levenshtein distance [18] to each unique word in the 1,164-word MacKenzie-Soukoreff phrase set [22] (our dictionary). The word with the smallest distance is selected as the recognized word.

4 EXPERIMENT

To evaluate the effectiveness of our system, we collected fingerspelling data from three participants and designed a personalized fingerspelling recognition model for each participant. P1 was a 25 year old Deaf male with native ASL fluency, P2 was a 19 year old hearing male with conversational ASL fluency, and P3 was a 18 year old female with intermediate ASL fluency. Following previous text-entry works [17], we selected the MacKenzie-Soukoreff Phrase Set [22], which consists of 500 phrases with 1,164 unique words. Each unique word was fingerspelled twice, resulting in a total of 2,328 words spelled by each participant. Data collection was carried out at the participant's leisure using a Google Pixel 4a smartphone with our push-to-sign app deployed. Participants were instructed to fingerspell words at a pace that felt most natural and comfortable to them.

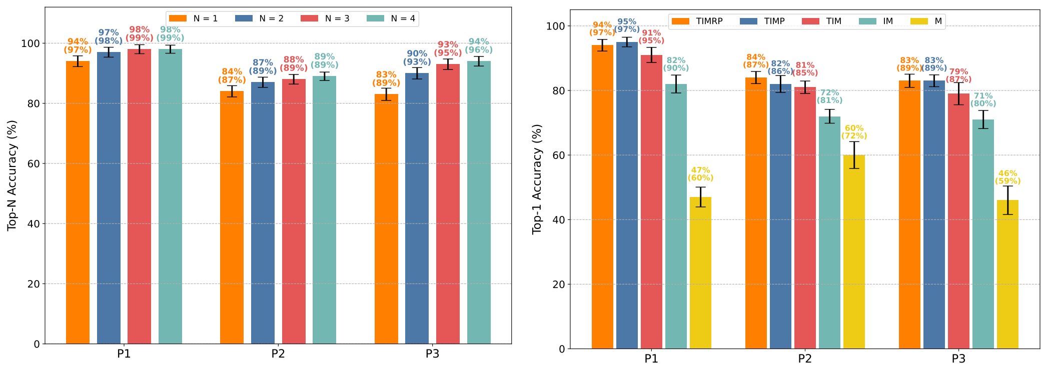
We evaluated our HMM-based backend with the collected dataset using a user-dependent 10-fold cross-validation protocol for each participant, where 10% of the data is reserved for testing and the remaining 90% is used for training in each fold. We report the average top-N (N=1,2,3,4) word and letter-level accuracies. We additionally conducted experiments to determine the ideal number of rings required for fingerspelling recognition. Specifically, we iteratively searched for the optimal configuration of the finger sensors by reducing the number of rings used for fingerspelling recognition. Starting with all five rings for evaluation, we removed a ring from each finger one-by-one to find best four-ring configuration. The four-ring configuration that showed least decrease in recognition accuracy was selected. We repeated this process until using a single ring for the evaluation. This greedy search approach finds an optimal number of ring sensors while maintaining recognition accuracy without trying all possible (5!) ring combinations.

4.1 Results

4.1.1 Top-N Word Recognition. Figure 2(a) shows the top-N (N = 1,2,3,4) word and character accuracy for each participant. The user-dependent model shows superior word accuracy on P1's data, achieving 94% as compared to P2's and P3's 83%. We attribute this variation among participants to their different fingerspelling speeds. P1, P2, and P3 fingerspelled at a pace of 27, 35, and 39 wpm, respectively. The rapid speed of fingerspelling in P2 and P3 can cause blurring in the transition between letters, posing a challenge for the model to accurately recognize individual letters.

Seeing that the top-4 accuracy greatly surpasses the top-1 accuracy, yielding a 7% increase in word accuracy , we consider a user interface that enables users to select words from the top 4 results in future fingerspelling live text entry systems has the potential to facilitate and support user interaction, similar to the word suggestion feature commonly found in mobile keyboards [29].

Finally, we ranked words based on their frequency of incorrect recognition. Repeated letters in words were found to be particularly prone to misrecognition. Across all participants, the top three words most frequently misrecognized were *steep, noon*, and *fell*, often misrecognized as *step, non*, and *feel*, respectively. The issue lies in how repeated letters are signed. Participants explained that they signed repeated letters by slightly shifting their hand to the right while maintaining the handshape of the letter, resulting in an



(a) (b)

Figure 2: (a) Top-N (N = 1,2,3,4) accuracy for each participant. Accuracy shown as word accuracy (character accuracy). (b) The top-1 accuracy for each participant when data from the best N-ring configuration (N = 1,2,3,4,5) are used for training and testing. Letters (T, I, M, R, P) denotes the fingers (thumb, index, middle, ring, pinky) that rings are worn.

unclear transition between repeated letters in the accelerometry signal that the recognizer struggles to classify.

4.1.2 Toward Smaller Form Factor. Figure 2(b) illustrates the top-1 word and character accuracy achieved by each participant when utilizing different N-ring configurations (N=1,2,3,4,5) for training and testing. We observed negligible change in performance upon removing the data collected from the ring worn on the ring finger. This indicates that the ring did not effectively capture meaningful information regarding fingerspelling. We further noticed that by solely utilizing data from the rings worn on the index and middle fingers, the model achieved a relative word accuracy of 85% – 88% and a relative character accuracy of 90% – 93% when compared to the model trained and tested on data collected from all 5 rings. This attributes to the fact that the index and middle fingers alone can capture the majority of variations in how letters are signed. The highly comparable performance observed in this experiment supports redesigning fingerspelling recognition systems, where users would only need to wear 2 rings instead of the 5 rings, which significantly reduces the burden on users in daily activities while maintaining a decent level of recognition accuracy.

5 CONCLUSION

We introduced FingerSpeller, a camera-free American Sign Language fingerspelling recognition system using smart rings. We assessed the system's performance by conducting tests on a dictionary comprising 1,164 words. Our system achieved an average character accuracy of 91% and word accuracy of 87% across three participants. Moreover, we demonstrated that the number of rings required for the system can be reduced from 5 to 2, while maintaining approximately 90% relative accuracy.

While our current system is primarily designed for seated text entry, it is crucial to explore the development of an on-the-go solution that enables users to perform text entry while engaged in other activities (e.g., walking). This expansion would greatly enhance usability and practicality of the system in everyday scenarios. Also, our system performs word-level recognition, which needs to be integrated to sentence-level delivery to enable live text entry to make conversations. One possible approach is to introduce a designated handshape and/or motion indicating spaces. By training a separate model to detect where the space gesture is signed, we can effectively segment the data and subsequently utilize it for continuous word recognition, thereby facilitating live text entry.

REFERENCES

[1] Manfred Georg Mark Sherwood Phil Culliton Sam Sepah Sohier Dane Thad Starner Ashley Chow, Glenn Cameron. 2023. Google - American Sign Language Fingerspelling Recognition. https://kaggle.com/competitions/asl-fingerspelling

[2] Benjamin J Bahan. 1996. *Non-manual realization of agreement in American Sign Language.* Boston University.

[3] Leonard E. Baum, Ted Petrie, George Soules, and Norman Weiss. 1970. A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *The Annals of Mathematical Statistics* (1970).

[4] Danielle Bragg, Oscar Koller, Mary Bellard, Larwan Berke, Patrick Boudreault, Annelies Braffort, Naomi Caselli, Matt Huenerfauth, Hernisa Kacorri, Tessa Verhoef, Christian Vogler, and Meredith Ringel Morris. 2019. Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility.*

[5] Ananya Choudhury, Anjan Kumar Talukdar, Kandarpa Kumar Sarma, and Manas Kamal Bhuyan. 2021. An adaptive thresholding-based movement epenthesis detection technique using hybrid feature set for continuous fingerspelling recognition. *SN Computer Science* 2 (2021), 1–21.

[6] Helen Cooper, Brian Holt, and Richard Bowden. 2011. Sign language recognition. *Visual Analysis of Humans: Looking at People* (2011), 539–562.

[7] Michael Erard. 2017. Why sign-language gloves don't help deaf people. *The Atlantic* (2017).

[8] Masaaki Fukumoto and Yoshinobu Tonomura. 1997. "Body coupled FingerRing" wireless wearable keyboard. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems.* 147–154.

[9] Kamala Gajurel, Cuncong Zhong, and Guanghui Wang. 2021. A fine-grained visual attention approach for fingerspelling recognition in the wild. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC).* IEEE, 3266–3271.

[10] Bogdan-Florin Gheran, Ovidiu-Ciprian Ungurean, and Radu-Daniel Vatavu. 2018. Toward Smart Rings as Assistive Devices for People with Motor Impairments: A Position Paper. In *RoCHI.* 99–106.

[11] Abraham Glasser, Vaishnavi Mande, and Matt Huenerfauth. 2021. Understanding Deaf and Hard-of-Hearing Users' Interest in Sign-Language Interaction with Personal-Assistant Devices. In *Proceedings of the 18th International Web for All Conference* (Ljubljana, Slovenia) *(W4A '21).* Association for Computing Machinery, New York, NY, USA, Article 24, 11 pages. https://doi.org/10.1145/3430263.3452428

[12] Abraham Glasser, Matthew Watkins, Kira Hart, Sooyeon Lee, and Matt Huenerfauth. 2022. Analyzing Deaf and Hard-of-Hearing Users' Behavior, Usage, and Interaction with a Personal Assistant Device That Understands Sign-Language Input. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing*

Systems (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 306, 12 pages. https://doi.org/10.1145/3491102.3501987

[13] Jun Gong, Zheer Xu, Qifan Guo, Teddy Seyed, Xiang 'Anthony' Chen, Xiaojun Bi, and Xing-Dong Yang. 2018. WrisText: One-Handed Text Entry on Smartwatch Using Wrist Gestures. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*.

[14] Joseph Hill. 2020. Do Deaf Communities Actually Want Sign Language Gloves? *Nature Electronics* 3, 9 (2020), 512–513. Issue 9.

[15] Yuya Igarashi, Kyosuke Futami, and Kazuya Murao. 2022. Silent Speech Eyewear Interface: Silent Speech Recognition Method Using Eyewear with Infrared Distance Sensors. In *Proceedings of the 2022 ACM International Symposium on Wearable Computers* (Cambridge, United Kingdom) *(ISWC '22)*. Association for Computing Machinery, New York, NY, USA, 33–38. https://doi.org/10.1145/3544794.3558458

[16] Lei Jing, Zixue Cheng, Yinghui Zhou, Junbo Wang, and Tongjun Huang. 2013. Magic ring: A self-contained gesture input device on finger. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia*. 1–4.

[17] Naoki Kimura, Tan Gemicioglu, Jonathan Womack, Richard Li, Yuhui Zhao, Abdelkareem Bedri, Zixiong Su, Alex Olwal, Jun Rekimoto, and Thad Starner. 2022. SilentSpeller: Towards Mobile, Hands-Free, Silent Speech Text Entry Using Electropalatography. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*.

[18] Vladimir I. Levenshtein. 1965. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics. Doklady* (1965).

[19] Hamzah Luqman, El-Sayed M El-Alfy, and Galal M BinMakhashen. 2021. Joint space representation and recognition of sign language fingerspelling using Gabor filter and convolutional neural network. *Multimedia Tools and Applications* 80 (2021), 10213–10234.

[20] Kent Lyons, Thad Starner, Daniel Plaisted, James Fusia, Amanda Lyons, Aaron Drew, and E. W. Looney. 2004. Twiddler Typing: One-Handed Chording Text Entry for Mobile Phones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.

[21] Kelly Mack, Danielle Bragg, Meredith Ringel Morris, Maarten W. Bos, Isabelle Albi, and Andrés Monroy-Hernández. 2020. Social App Accessibility for Deaf Signers. *Proc. ACM Hum.-Comput. Interact.* (2020).

[22] I. Scott MacKenzie and R. William Soukoreff. 2003. Phrase Sets for Evaluating Text Entry Techniques. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*.

[23] Ross Mitchell, Travas Young, Bellamie Bachleda, and Michael Karchmer. 2006. How Many People Use ASL in the United States? Why Estimates Need Updating. *Sign Language Studies* (2006).

[24] Masa Ogata, Yuta Sugiura, Hirotaka Osawa, and Michita Imai. 2012. IRing: Intelligent Ring Using Infrared Reflection. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology* (Cambridge, Massachusetts, USA) *(UIST '12)*. Association for Computing Machinery, New York, NY, USA, 131–136. https://doi.org/10.1145/2380116.2380135

[25] Carol A Padden and Darline Clark Gunsauls. 2003. How the alphabet came to be used in a sign language. *Sign Language Studies* (2003), 10–33.

[26] Kseniia Palin, Anna Maria Feit, Sunjun Kim, Per Ola Kristensson, and Antti Oulasvirta. 2019. How do people type on mobile devices? Observations from a study with 37,000 volunteers. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–12.

[27] Katerina Papadimitriou and Gerasimos Potamianos. 2019. End-to-End Convolutional Sequence Learning for ASL Fingerspelling Recognition.. In *INTERSPEECH*, Vol. 2019. 2315–9.

[28] David Quinto-Pozos. 2010. Rates of fingerspelling in american sign language. In *Poster presented at 10th Theoretical Issues in Sign Language Research conference, West Lafayette, Indiana*, Vol. 30.

[29] Quentin Roy, Sébastien Berlioux, Géry Casiez, and Daniel Vogel. 2021. Typing Efficiency and Suggestion Accuracy Influence the Benefits and Adoption of Word Suggestions. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*.

[30] Shikhar Sharma and Krishan Kumar. 2021. ASL-3DCNN: American sign language recognition technique using 3-D convolutional neural networks. *Multimedia Tools and Applications* 80, 17 (2021), 26319–26331.

[31] B. Shi, A. Martinez Del Rio, J. Keane, D. Brentari, G. Shakhnarovich, and K. Livescu. 2019. Fingerspelling recognition in the wild with iterative visual attention. *ICCV* (October 2019).

[32] Tanmay Srivastava, Prerna Khanna, Shijia Pan, Phuc Nguyen, and Shubham Jain. 2022. MuteIt: Jaw Motion Based Unvoiced Command Recognition Using Earable. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* (2022).

[33] Zixiong Su, Shitao Fang, and Jun Rekimoto. 2023. LipLearner: Customizable Silent Speech Interactions on Mobile Devices. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 696, 21 pages. https://doi.org/10.1145/3544548.3581465

[34] Tap. 2019. Tap Strap 2. https://www.tapwithus.com/product/tap-strap-2/.

[35] Radu-Daniel Vatavu and Ovidiu-Ciprian Ungurean. 2022. Understanding Gesture Input Articulation with Upper-Body Wearables for Users with Upper-Body Motor Impairments. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–16.

[36] Mathias Wilhelm, Daniel Krakowczyk, Frank Trollmann, and Sahin Albayrak. 2015. eRing: multiple finger gesture recognition with one ring using an electric field. In *Proceedings of the 2nd international Workshop on Sensor-based Activity Recognition and Interaction*. 1–6.

[37] Steve J. Young, Gunnar Evermann, Mark John Francis Gales, Thomas Hain, Dan J. Kershaw, Gareth L. Moore, J. J. Odell, David Ollason, Daniel Povey, Valtchev, and Philip C. Woodland. 2006. The HTK book version 3.4. Cambridge Univ. Eng. Dept. Speech Group and Entropic Research Lab. Inc.

[38] Cheng Zhang, Anandghan Waghmare, Pranav Kundra, Yiming Pu, Scott Gilliland, Thomas Ploetz, Thad E. Starner, Omer T. Inan, and Gregory D. Abowd. 2017. FingerSound: Recognizing Unistroke Thumb Gestures Using a Ring. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 120 (sep 2017), 19 pages. https://doi.org/10.1145/3130985

[39] Cheng Zhang, Xiaoxuan Wang, Anandghan Waghmare, Sumeet Jain, Thomas Ploetz, Omer T. Inan, Thad E. Starner, and Gregory D. Abowd. 2017. FingOrbits: Interaction with Wearables Using Synchronized Thumb Movements. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers* (Maui, Hawaii) *(ISWC '17)*. Association for Computing Machinery, New York, NY, USA, 62–65. https://doi.org/10.1145/3123021.3123041

[40] Hao Zhou, Taiting Lu, Yilin Liu, Shijia Zhang, and Mahanth Gowda. 2022. Learning on the Rings: Self-Supervised 3D Finger Motion Tracking Using Wearable Sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* (2022).