APPLIED SCIENCES AND ENGINEERING

An AI-driven, wearable, conformal ring system for real-time and user-independent sign language interpretation

Jaejin Park[1,2]†, Yejee Shin[1]†, In Sik Min[1], Yeeun Lee[3], Jiwon Lee[1], Jung-Hoon Hong[1], Sunho Park[4], Sang Hoon Park[1], Junhyeong Baek[1], Taemin Kim[5,6], Kyubeen Kim[7], Doyun Kim[8], Yunkyu Park[9], Jongyoo Kim[1], Young Uk Cho[10], Yeonsik Choi[4], Byunghwan Jeon[11], Kyowon Kang[12]*, Dosik Hwang[1,13,14,15]*, Ki Jun Yu[1,16]*

Sign language translation systems have long aimed to bridge the communication between signers and nonsigners. However, preliminary systems rely on glove-type wearables or wired sensor arrays, which constrain hand movement, reduce comfort, and require nonpersonalized sensor positions that limit adaptability across users. Here, we introduce a wirelessly connected, ring-type sign language translator (WRSLT) designed to overcome these limitations by enabling full finger mobility through independent sensor rings and multilink communication. The system supports static and dynamic gesture detection using selected fingers via quantitative relevance analysis and achieves robust user-independent performance without per-user calibration. WRSLT demonstrated high recognition accuracy on large-scale datasets comprising 100 American Sign Language and 100 International Sign Language words, achieving 88.3 and 88.5% accuracy, respectively, under unseen-user conditions (i.e., test users not included in model training). Furthermore, a custom sequential word detection framework enables sentence-level translation from continuous signing input without requiring separate training on entire sentence structures.

INTRODUCTION

Sign language is a primary communication method for individuals with hearing or speech impairments, conveying meaning through hand gestures and body movements (*1*, *2*). However, individuals without prior knowledge of sign language are typically unable to understand or respond, resulting in a consistent communication barrier with signers in daily interactions. In addition, more than 300 different sign languages are used worldwide, each developed based on regional and cultural contexts (*3*). These variations also make it difficult for signers from different regions to communicate with one another. To promote more effective communication for sign language users, a range of translation device has been developed to recognize sign language.

Various technological approaches have been proposed to translate sign language into spoken or written language, which can be broadly categorized into vision-based and wearable sensor-based systems. Vision-based systems typically rely on external cameras combined with computer vision algorithms to recognize hand gestures and motion trajectories (*4*–*6*). While these systems can achieve reasonable performance under controlled environments (i.e., fixed camera position), their applicability is often limited to predefined settings where the camera can consistently capture the signer's hand movements. Moreover, they are sensitive to occlusion, lighting variations, and background complexity, making it difficult to accurately recognize dynamic gestures in real-world scenarios. In parallel, wearable sign language translators have been developed using different types of body-attached sensors to directly measure physical or electrophysiological signals associated with hand gestures. These systems typically integrate various sensor types, such as surface electromyography (EMG) electrodes (*7*–*9*), strain and bending sensors (*10*–*12*), triboelectric sensors (*13*, *14*), and inertial measurement units (*15*–*17*). Sensors are attached to the skin or embedded into wearable structures, allowing for direct and continuous measurement of hand posture and movement without being affected by lighting, occlusion, or background interference.

Sign language recognition systems were primarily on the basis of wired configurations, where each sensor was physically connected to an external processing unit via data transmission lines. While this approach enabled basic signal acquisition, the constant presence of cables between sensors and the processing unit substantially restricted hand movement and limited usability in real-world environments. For the mobility and usability, recent studies adopted wireless architectures that eliminate the need for direct cabling between the sensors and the external processing unit, allowing for improved wearability

[1]Department of Electrical and Electronic Engineering, Yonsei University, 50, Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea. [2]George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA. [3]Department of Artificial Intelligence, Yonsei University, 50, Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea. [4]Department of Materials Science and Engineering, Yonsei University, 50, Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea. [5]Department of Electrical and Computer Engineering, Ajou University, Suwon 16499, South Korea. [6]Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA 22904, USA. [7]Department of Mechanical Engineering and Materials Science, Washington University in St. Louis, St. Louis, MO 63105, USA. [8]Department of Chemistry, Hankuk University of Foreign Studies (HUFS), Yongin 17035, Republic of Korea. [9]Department of Physics, Memory and Catalyst Research Center, Hankuk University of Foreign Studies, Yongin 17035 Republic of Korea. [10]Department of Biomedical & Robotics Engineering, Incheon National University, Yeonsu-gu, Incheon 22012, Republic of Korea. [11]Division of Computer Engineering Hankuk University of Foreign Studies, Yongin 17035, South Korea. [12]Department of Biomedical Engineering, Hankuk University of Foreign Studies, Yongin 17035, Republic of Korea. [13]Department of Radiology and Research Institute of Radiological Science and Center for Clinical Imaging Data Science, Yonsei University College of Medicine, Seoul, Republic of Korea. [14]Artificial Intelligence and Robotics Institute, Korea Institute of Science and Technology, Seoul, Republic of Korea. [15]Department of Oral and Maxillofacial Radiology, Yonsei University College of Dentistry, Seoul, Republic of Korea. [16]Department of Electrical and Electronic Engineering, YU-Korea Institute of Science and Technology (KIST) Institute, Yonsei University, Seoul 03722, Republic of Korea.
*Corresponding author. Email: kijunyu@yonsei.ac.kr (K.J.Y.); dosik.hwang@yonsei.ac.kr (D.H.); kyowonkang@hufs.ac.kr (K.K.)
†These authors contributed equally to this work.

and broader application in mobile or real-life environments (*18*–*21*). However, despite the use of wireless communication modules, a major limitation still remains in the physical wiring between sensors within the device itself. In typical designs, multiple sensors such as inertial or strain sensors must be positioned across different fingers to capture detailed motion data. These sensors are often interconnected through physical data transfer lines that route signals to a single shared wireless data transmitter, forming what is essentially a partially wired system. This structure often causes physical discomfort and limits natural hand movement due to the presence of numerous interconnecting wires. It also interferes with the execution of complex and dynamic gestures required for accurate sign language translation, reducing the system's practicality and accuracy for real-world use.

In preliminary studies, wearable sign language translation systems have been implemented as smart gloves (*22*–*27*). This form factor offers a convenient and well-established platform for integrating multiple sensors across the hand, enabling comprehensive data collection from multiple joints and fingers within a single garment. However, despite these advantages, glove-based systems present several limitations in terms of comfortability and adaptability. The enclosed nature of gloves reduces breathability, making prolonged use uncomfortable, especially during continuous or intensive long-term signing. Moreover, glove-based systems typically assume fixed sensor placements on the glove, which do not account for individual variations in hand size, finger length, or joint positions. As a result, the sensor alignment may deviate from the intended anatomical locations, leading to inconsistent signal acquisition and reduced recognition accuracy across users.

In parallel with these hardware-based developments, recent advances in sign language recognition have emerged through multidisciplinary approaches that combine wearable sensor technologies with artificial intelligence (AI)–based algorithms by capturing and interpreting complex gesture patterns (*28*–*32*). The development of deep learning models capable of generalizing across diverse signers has become increasingly important for practical applications. In particular, ensuring high recognition accuracy not only when the test user is included in the training data (seen-user condition) but also when the system is applied to general users (unseen-user condition; test user is not included in the training data) is important. Achieving robust performance in unseen-user condition is essential for developing user-independent devices and is a prerequisite for broad deployment in practical settings. Also, recent efforts have extended beyond isolated word recognition toward continuous, sentence-level translation (*13*). This allows sequential hand gestures to be interpreted as temporally connected patterns rather than discrete, independent signs. This functionality is particularly critical for practical deployment, as sign language is typically conveyed through fluid and uninterrupted gesture sequences in real-world settings.

Here, we introduce a wirelessly connected, ring-type sign language translator (WRSLT) that integrates practical hardware design and deep learning architecture. Unlike conventional glove-based designs, where sensors are preembedded at fixed positions within the glove fabric, the proposed system is ring-type sensors that can be independently positioned on each finger. This modular configuration allows flexible sensor positioning that naturally adapts to variations in hand size, joint alignment, and finger morphology across users, enhancing both comfort and signal reliability. To minimize hardware complexity without compromising recognition performance, we quantitatively evaluated the contribution weight of each finger to overall classification accuracy. On the basis of this analysis, seven dominant fingers were identified, enabling optimized sensor allocation. The WRSLT uses a fully wireless architecture in which each ring operates independently without physical wiring between sensor modules, thereby enabling unrestricted and seamless hand movement during dynamic signing, whereas conventional systems rely on centralized, wired connections that constrain natural motion. The system captures gravity-oriented hand gesture feature using accelerometer sensors rather than relying on biosignals such as EMG, which are highly user-specific and require extensive training (table S1). In contrast to vision-based approaches or human interpretation, these strategies enables continuous, privacy-preserving and environment-independent operation without reliance on external infrastructure or human availability. Building on the motion derived gravity-based signals, we propose a deep learning architecture that leverages pretrained foundation models to enhance recognition capabilities. By extracting hand movement features during signing from these foundation models and defining them as prototypes for recognition, our approach focuses on not only for users included in the training dataset but also for unseen users. This generalizability and user independency without requiring user-specific adaptation enables high recognition accuracy even for individuals who did not participate in the training phase (unseen-user condition). In addition, the system supports sentence-level translation and is compatible with multiple sign languages including American Sign Language (ASL) and International Sign Language (ISL), which demonstrates worldwide appliable property. Sentence-level translation is enabled by a sequential gesture processing framework, referred to as sequential word detection framework, which detects and classifies continuous input streams without requiring manual segmentation of the sentence. These combined features make the proposed system highly adaptable, user-independent, multisign language recognizable and suitable for practical deployment in diverse communication scenarios. In contrast to conventional glove type or partially wired systems restricted to word-level recognition and user specific tuning, the WRSLT enables real-time, sentence-level translation using a fully wireless and modular architecture that generalizes across users and sign languages.

RESULTS

Characterization of WRSLT

Figure 1A illustrates the overall concept of WRSLT. The device is worn on the fingers of a signer, and hand gestures are captured and translated into corresponding text-based outputs in real time. Unlike word-level recognition systems, the WRSLT enables sentence-level translation by identifying and segmenting sequences of trained words from continuous signing input. Furthermore, the system accommodates multiple sign languages, such as ASL and ISL, by independently training models tailored to each language. Figure 1B presents the structural configuration of a single ring device, which comprises a three-axis accelerometer, power management unit, Bluetooth low energy (BLE) system-on-chip (SoC), elastic ring platform, and magnet. Detailed circuit schematic and circuit components are shown in fig. S1. The device also includes a compact battery that lasts nearly 12 hours (fig. S2) can be easily replaced, enabling continuous use without full system disassembly. The device is designed to measure both angular orientation and dynamic motion of each finger through acceleration signals. Similar to the daily wearing rings, WRSLT is strategically positioned on the proximal phalanx, the segment of the

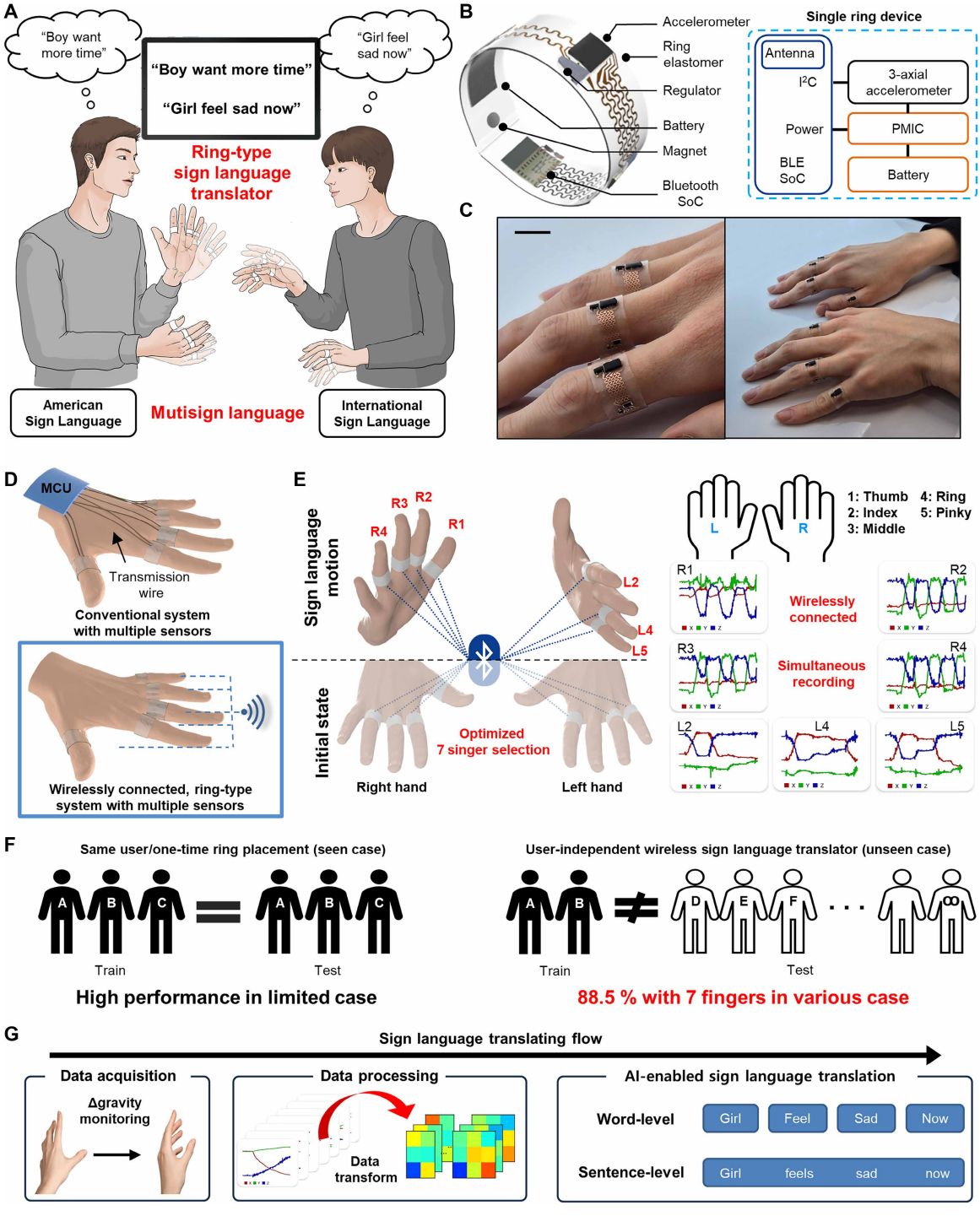

Fig. 1. Overall concept and design of WRSLT. (**A**) Conceptual illustration of WRSLT enabling intuitive sentence-level translation and compatibility with multiple sign languages (e.g., ASL and ISL). (**B**) Structural diagram and circuit layout of a single WRSLT which is fabricated on a flexible substrate, allowing it to be bent into a ring shape for finger mounted application. (**C**) Optical images of WRSLT worn on the fingers with the enlarged image. Scale bar, 1 cm. (**D**) Comparison of conventional multi-sensor systems and the proposed WRSLT architecture. Traditional wired setups rely on physical transmission wires connecting each sensor to a centralized microcontroller unit (MCU), limiting hand mobility. In contrast, WRSLT adopts a wirelessly connected design that eliminates interconnecting cables, allowing for unrestricted and natural hand movement. (**E**) Schematic illustration of the WRSLT network using Bluetooth multilink communication to enable simultaneous data acquisition across selected fingers. (R: right hand, L: left hand, 1: thumb, 2: index finger, 3: middle finger, 4: ring finger, 5: little finger). (**F**) Comparison between seen-user and unseen-user scenarios in sign language recognition. The unseen-user scenario evaluates generalizability by testing on users not included in the training set, demonstrating user-independent performance. (**G**) Overall system flow of the WRSLT platform for real-time sign language translation which supports both word-level and sentence-level translation. [(D), (E), and (G)] Hand model was used with permission from Artec 3D (https://www.artec3d.com), CC BY 4.0 https://creativecommons.org/licenses/by/4.0/deed.en.

finger located between the proximal interphalangeal (PIP) joint, and the metacarpophalangeal (MCP) joint (*33*). The interphalangeal joints, such as PIP, allow for rotational motion in the coronal plane, facilitating flexion and extension. Meanwhile, MCP joints support motion in both the coronal and sagittal planes, enabling not only flexion and extension but also side-to-side movements such as abduction and adduction. This anatomical placement enables the system to capture a wide spectrum of finger movements with high fidelity. The accelerometer continuously captures motion signals, while the onboard circuitry manages power regulation and enables wireless data transmission via the BLE module. To accommodate repeated finger bending during sign gestures, serpentine-structured interconnections are incorporated to improve mechanical stability and maintain electrical integrity under dynamic deformation (fig. S3 and movie S1). This self-contained design enables each ring to operate independently without requiring physical interconnections between sensors. The photographs of the WRSLT worn on the fingers are shown in Fig. 1C, demonstrating their lightweight, flexible, and skin-conformal form factor designed for stable operation with magnetic fixation ensuring secure yet detachable placement on the finger (movie S2). Finger circumference measurements across multiple users are provided in fig. S4, showing that the device conforms securely around the finger without inducing excessive strain on the copper interconnects (fig. S5), thus ensuring mechanical safety during repeated bending motions. The intended placement on the proximal phalanges is visible during actual use, while fig. S6 presents the device appearance before mounting.

A key distinction between conventional multisensor systems and the proposed WRSLT platform is the implementation of a wirelessly connected architecture (Fig. 1D). In conventional systems, multiple sensors are physically connected to a centralized microcontroller unit (MCU) via transmission wires, and the MCU handles data processing or wireless communication by transmitting the aggregated data to an external receiver. However, this approach not only adds physical bulk but also introduces substantial mechanical constraints due to the increasing number of transmission wires, especially as more sensors are added. These wired configurations notably restrict hand mobility, which is essential for accurately capturing the dynamic and continuous gestures involved in sign language translation. In contrast, the WRSLT is composed of compact, independently operated ring-type systems, each equipped with its own wireless transmission module. This decentralized architecture allows the system to eliminate data transmission lines entirely, enabling unrestricted finger motion and supporting dynamic gesture recognition with minimal mechanical constraint.

Wirelessly connected architecture is achieved through Bluetooth multilink technology, allowing multiple sensors to operate simultaneously without physical connections (Fig. 1E). Each ring establishes an independent BLE peripheral link to the host device, enabling concurrent transmission of motion data from multiple fingers. Because every ring module performs its own sensing and communication, signals are acquired in parallel rather than sequentially, and the host device aggregates the incoming packets in the order of reception, providing multifinger data streams at the application level. This approach eliminates the need for any physical interconnections or shared wiring between sensors. The wireless multinode structure not only preserves full finger mobility but also ensures mechanical reliability, as each ring is mounted on a small and curved region where a simplified, compact circuit is essential for stable operation. By enabling independent measurement of each finger, the system supports quantitative evaluation of the contribution of individual fingers to sign language translation. On the basis of this analysis, seven fingers were identified as most dominant in classification performance, allowing for number of sensor reductions without compromising accuracy (Fig. 1E). The selected fingers—R1 (right thumb), R2 (right index), R3 (right middle), R4 (right ring), L2 (left index), L4 (left ring), and L5 (left little)—were integrated into the final system and wirelessly connected for simultaneous data recording.

Figure 1F highlights the distinction between seen-user and unseen-user scenarios in evaluating the performance of sign language translation systems. In AI-based gesture recognition, ensuring generalizability and user independency requires that training and testing be conducted on independent datasets. When the same user participate in both training and testing (seen-user condition)—especially under identical recording conditions, such as one-day acquisition with a single attachment without sensor repositioning—the model may overfit to user-specific patterns. This often leads to inflated accuracy in seen-user evaluations while significantly degrading performance in unseen-user scenarios where robust generalization is essential. However, in practical applications, gestures can vary subtly across users, and sensor reattachment may cause slight shifts in placement, even for the same individual. These variations inherently reduce accuracy, particularly under unseen-user conditions where trainees are not included in test data. To validate the generalizability of the system, the WRSLT was evaluated under an unseen-user setting, in which the test users did not participate in training. Despite interuser variability and potential differences in sensor alignment, the system achieved high recognition accuracies of 88.4% for 100 words of ASL and 88.9% for 100 words of ISL, using only seven fingers. These results demonstrate the broad applicability of user-independent sign language translation, validating the generalizability and robustness of the proposed platform across different users and usage conditions.

The overall flow of sign language translation using the WRSLT is presented in Fig. 1G, from sensor-level data acquisition to AI-driven linguistic output. As hand gestures are performed, each ring device captures gravity-oriented motion signals through embedded inertial sensors. These signals are continuously collected and converted into time-series representations suitable for classification by pretrained deep learning models. For word-level recognition, each gesture is processed individually to match with trained gesture templates. In addition to isolated word detection, the WRSLT also enables sentence-level translation by applying a sequential word detection framework that identifies and groups consecutive gestures into meaningful sequences. Specifically, a sliding window–based approach is used to track transitions across time and assemble a structured sentence output. This sequential analysis reflects the fluid nature of real-life signing, where expressions are typically conveyed as continuous gesture streams rather than isolated words. The ability to translate at both the word and sentence levels enhances the system's practicality and applicability in real-world communication scenarios.

Hand gesture detection using acceleration

Among various wearable sensing modalities, we adopted a gravity-based inertial sensing strategy because it offers stable, relatively invariant gesture features compared to electrophysiological or strain responses, which are essential for user-independent sign language recognition. EMG systems exhibit large interuser variability due to differences in muscle physiology, electrode-skin impedance, and placement sensitivity, making per-user calibration unavoidable in practice.

Strain-based sensors can capture bending-induced deformation but cannot provide absolute finger orientation or tilt information, which is critical for distinguishing many static sign postures. In contrast, a three-axis accelerometer directly measures both the projection of gravity under static conditions and motion-induced accelerations during dynamic transitions. This enables simultaneous acquisition of static postural information and dynamic gesture trajectories from a single compact sensor module.

Therefore, a three-axis accelerometer is integrated into each ring to sense finger motion. The principle of tilt estimation using the onboard accelerometer embedded in the WRSLT is illustrated in Fig. 2A. The accelerometer measures acceleration in its local coordinate system (x', y', z'), which shifts in orientation relative to the horizontal coordinate frame (x, y, z) as the device is tilted. Under static conditions, the only measurable acceleration is due to gravity (1 g), which serves as a reference vector. By comparing the gravity vector's projection onto the accelerometer's axes, the relative tilt of the device can be inferred. This enables continuous monitoring of finger orientation by capturing deviations between the sensor's internal reference frame and the external horizontal frame. Figure 2B demonstrates the static tilt response of the WRSLT under controlled angular displacements. "Static" refers to quasi-static tilt steps, where the device is rotated to a target angle and held stationary to capture the steady-state gravitational projection. The system was placed on a horizontal surface and tilted sequentially to 30°, 60°, and 90°, each repeated twice in a single axis direction. As expected, the projected gravitational component along the sensing axes changes with tilt angle, particularly in the x- and z-axis signals. These variations follow the sine of the tilt angle due to the directional projection of gravity. For example, the gravitational force component along the horizontal axis is ~0.5 g for 30° (sin 30°) and ~0.87 g for 60° (sin 60°). At 90°, the axis aligned with gravity measures close to 1 g, while orthogonal axes approach zero. The y axis remains largely unaffected, consistent with the planar rotation. On the basis of these gravity-oriented measurements, pitch (rotation about the x axis) and roll (rotation about the y axis) angles can be analytically estimated using trigonometric relationships derived from accelerometer readings, as shown in fig. S7. Figure 2C presents the dynamic response of the accelerometer during rotational motion at three different angular velocities: 20°/s, 33.3°/s, and 100°/s. "Dynamic" refers to continuous, nonstationary rotation, where the device is moved through the angular range without holding at intermediate angles. Distinct differences in waveform frequency and slope are observed across the speed regimes, highlighting the system's capability to differentiate gesture velocity through temporal changes in acceleration magnitude. These results confirm the WRSLT's effectiveness in capturing both static orientation and dynamic motion features, which are essential for accurate recognition of both posture-based and motion-based gestures in sign language. In addition, the accelerometer is consistently positioned on the proximal phalanx for gesture detection, and the baseline variation induced by repetitive attachment and detachment was negligible (<0.3%), demonstrating robust sensing stability during daily use (fig. S8).

Figure 2D presents representative dynamic gesture data obtained during real-time signing using the WRSLT. Three commonly used sign language words (want, good, and animal) are selected to demonstrate various motion patterns. The accelerometer coordinates of the R2 sensor are visualized alongside corresponding gesture images to provide a spatial reference for hand orientation during signing. Similar waveform patterns were observed across multiple fingers within the same hand, particularly among R2, R3, and R4 on the right hand and L2, L4, and L5 on the left. This consistency indicates that the WRSLT provides stable and reproducible measurements for fingers with similar anatomical orientation. In contrast, the R1 sensor on the thumb showed a different pattern due to its naturally angled resting position, which differs from the alignment of the other fingers. This variation highlights the system's ability to detect finger-specific motion patterns, even when anatomical orientations differ, such as in the case of the thumb.

During the "want" gesture, both hands transition from an open-palm posture to inward-facing fists. Initially, the phalanges are aligned such that the z axis of the accelerometer faces downward, resulting in a gravity-induced reading near $-1g$. As the hands move inward to form fists, the x axis of the accelerometer rotates downward, producing a $-1g$ component on the x axis, while the z axis rotates forward, approaching $0g$. This gesture is therefore characterized by bilateral, time-varying acceleration profiles with clear differences between initial and final states. In contrast, the "good" gesture involves asymmetric hand movement: The left hand remains stationary while the right hand moves from the lips to the left palm. As the z axis of the accelerometer in R2 transitions from a horizontal to downward-facing orientation, it exhibits a marked change in gravitational component—from near-zero at the start to ~$-1g$ at the end. The stationary left hand produces near-flat signals, demonstrating the ability of WRSLT to distinguish between dynamic and static hand segments. The "animal" gesture features a symmetrical and repetitive inward motion of both fists. As the gesture begins and ends with the same posture, the initial and final acceleration values are nearly identical. However, dynamic transitions in between generate measurable fluctuations, especially along the z axis, which reflects not only gravitational projection but also motion-induced acceleration. These results demonstrate that WRSLT effectively captures both static postural states and dynamic gesture transitions, enabling robust recognition of a wide range of sign language components.

Algorithms for sign language translation

Our goal is to classify total 200 sign language words (100 ASL + 100 ISL) using gravity-oriented hand gesture signals captured with 50 sampling rate. To support effective communication, we focused on 100 frequently used words that commonly appear in daily conversations. This vocabulary size enables the construction of basic sentence structures and supports sufficient interactions in everyday communication scenarios. To align the gesture signals with their corresponding semantic meaning, we adopt a multimodal neural network framework that incorporates both a pretrained text encoder and a custom-designed signal encoder. The detailed architecture is illustrated in Fig. 3A, and the corresponding layer-wise parameters are summarized in fig. S9. The text encoder leverages a pretrained foundation model to generate rich semantic embeddings of each target word (*34*). Raw sensor data were preprocessed by mean shifting and temporal smoothing to reduce noise before convolutional feature extraction. To address the fine-grained differences between semantically distinct but visually similar gestures, we used a quantization module that transforms continuous sensor inputs into 5 discrete levels based on predefined thresholds. The quantization module discretized continuous sensor inputs into five levels using predefined thresholds: values greater than or equal to 105 were mapped to 0, values between 61 and 104 to 1, between −60 and 60 to 2, between −104 and −61 to 3, and values less than −104 to 4. These quantized features were passed

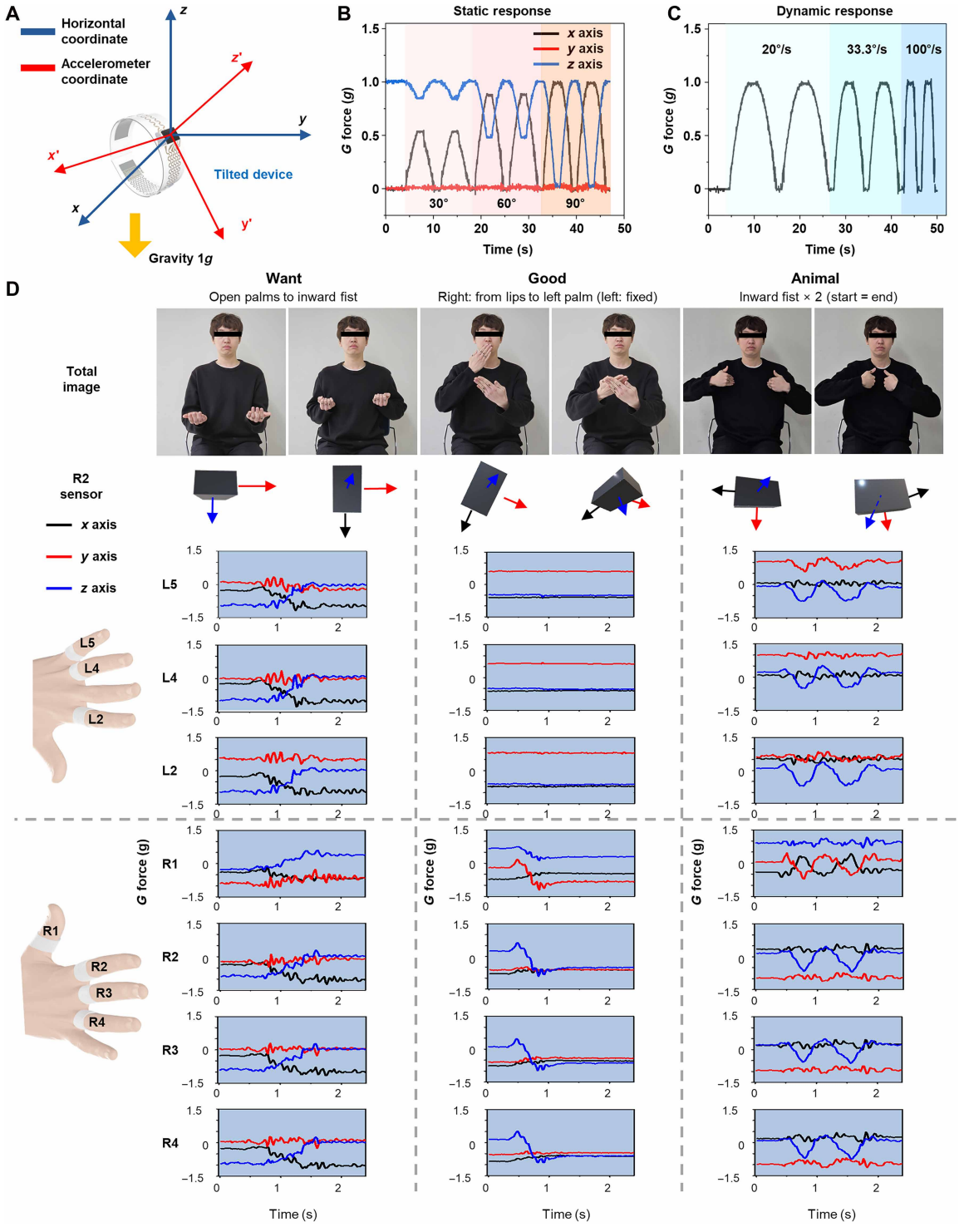

Fig. 2. Characterization of gravity-oriented sensing using a single WRSLT. (**A**) Schematic of the accelerometer's local coordinate frame (x', y', z') relative to the horizontal frame (x, y, z), showing the tilt-dependent projection of the gravity vector. (**B**) Static tilt measurements demonstrating the *G* force response along each axis as the device is tilted to 30°, 60°, and 90° on a flat surface. The measured acceleration closely follows the sine of the tilt angle. (**C**) Dynamic response of the system to repeated tilting motions at three different angular velocities (20°/s, 33.3°/s, and 100°/s). (**D**) Representative gesture data for "want," "good," and "animal" captured from selected fingers using WRSLT. Hand model was used with permission from Artec 3D (https://www.artec3d.com), CC BY 4.0 https://creativecommons.org/licenses/by/4.0/deed.en.

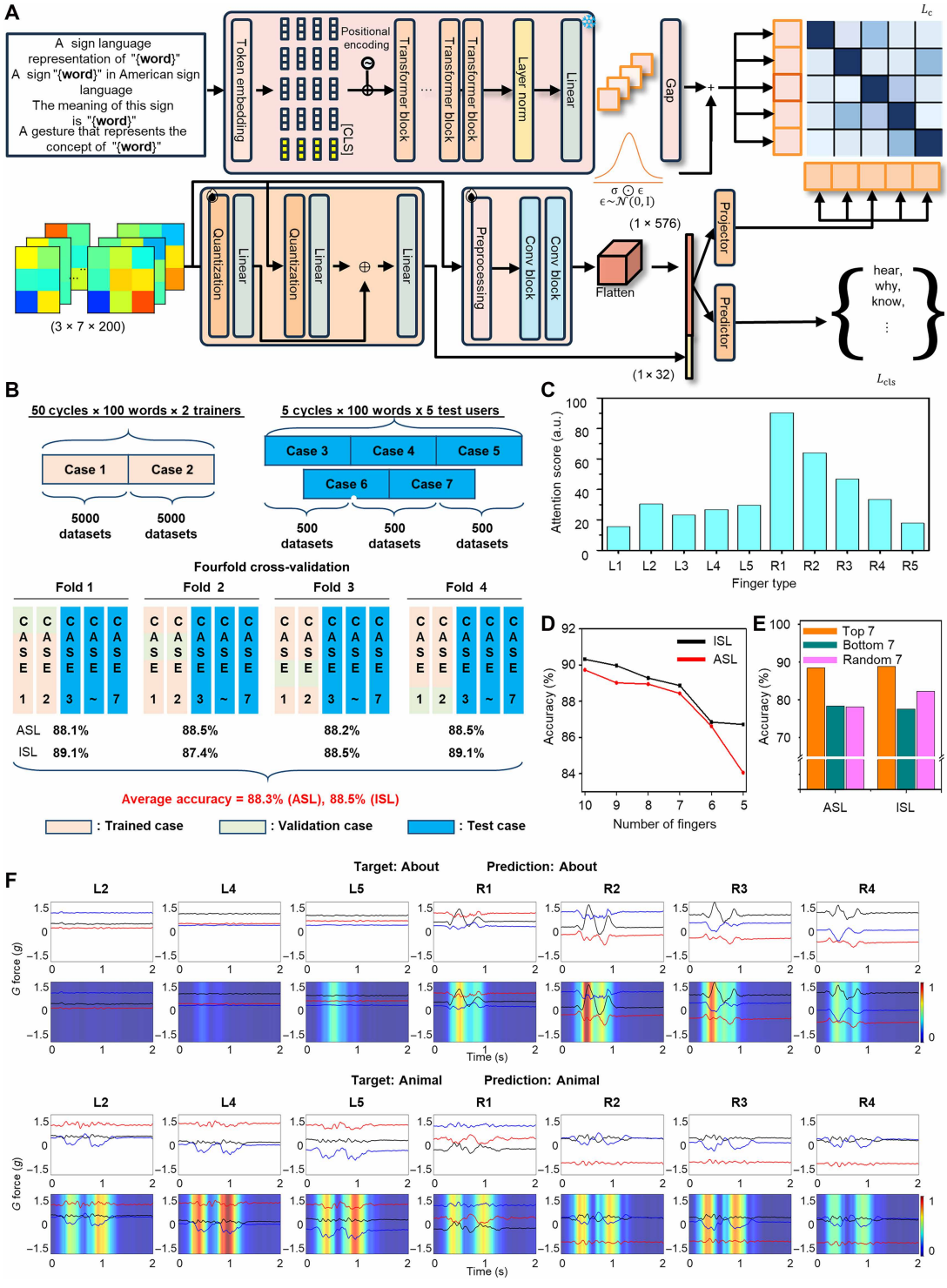

Fig. 3. Algorithmic architecture and detailed performance of WRSLT. (**A**) Schematic of multimodal neural network architecture combining a pretrained text encoder and a custom-designed sensor encoder for sign language classification. (**B**) Dataset configuration and fourfold cross-validation strategy. In each fold, data from two users are used for training and validation, and the model is evaluated on three unseen users. (**C**) Finger-wise attention score distribution, identifying dominant contributors to classification performance. (**D**) Recognition accuracy as a function of the number of fingers used, highlighting performance saturation at seven sensors. (**E**) Accuracy comparison between top seven and bottom seven finger sets for both ASL and ISL datasets. (**F**) Relevance-weighted class activation maps (R-CAMs) for two representative words. Warmer colors (e.g., red/yellow) indicate high relevance, while cooler colors (blue) indicate low relevance.

through a linear layer to produce features. To create semantic targets, embeddings from four distinct natural language prompts per word were averaged to form class prototypes. These prototypes act as anchors in the shared latent space, promoting a semantically structured embedding landscape. This design encourages the model to project gesture signals that are visually similar but semantically unrelated into distinct regions of the latent space, thereby improving its ability to resolve fine-grained interclass ambiguities. The signal encoder combined features from the quantization and convolutional paths, projecting the concatenated features into a shared latent space aligned with the text embeddings. The model was trained using a dual loss function combining cross-entropy for classification and cosine similarity to encourage semantic alignment, with Gaussian noise added to text embeddings to improve generalization.

The overall evaluation setup is demonstrated in Fig. 3B. The dataset was constructed using seven wirelessly connected ring-type sensors placed on selected fingers (R1 to R4, L2, L4, and L5), with simultaneous bilateral recording. Two sign languages were included in the study: ASL and ISL, each comprising 100 words. For both ASL and ISL, we constructed datasets with 5000 samples per experimental case and conducted fourfold cross-validation to evaluate user-independent performance. In each fold, five unseen participants were assigned to the test set. For ISL, the results shown in Fig. 3B show an average accuracy of 88.5% across folds. For ASL, the average accuracies per fold were 88.1, 88.5, 88.2, and 88.5%, yielding an overall mean of 88.3%.

The importance of each finger in sign classification was quantitatively assessed using layer-wise relevance propagation (LRP) (*35*), as illustrated in Fig. 3C, to make a balance between recognition performance and user comfort by reducing the number of required sensing channels by maintaining high accuracy. Attention scores derived from the LRP analysis indicate that certain fingers consistently contribute more to gesture recognition than others. The total attention score of each finger across all folds is summarized in fig. S10. Notably, R1 (right thumb) exhibits the highest relevance score among all fingers. This prominence is attributed to both anatomical and behavior factors. In sign language, gestures are typically performed with the dominant hand, which in most datasets—including ours—is the right hand. As a result, fingers on the right hand often show higher involvement in articulation. Furthermore, sensors are placed on the proximal phalanges of each finger, and while index through pinky fingers generally share similar orientations, the thumb rests at an anatomically distinct angle. This oblique orientation of R1 leads to a unique sensing direction, enabling it to capture movement features that are not redundant with the other fingers. Some sign gestures involve whole-hand motions rather than isolated finger movements, resulting in similar signal patterns across R2 to R5. In these cases, R1's distinct orientation plays a key role in distinguishing gestures by providing nonoverlapping spatial information, thereby contributing disproportionately to recognition accuracy. On the basis of these relevance scores, Fig. 3D and fig. S11 shows the classification performance achieved using different numbers of fingers. The results demonstrate that high recognition accuracy can be maintained with as few as 7 to 10 fingers, confirming that performance saturation is reached without requiring full-hand instrumentation. To compare translation accuracy based on finger relevance, Fig. 3E presents results using the top seven most relevant fingers against the bottom seven and the randomly selected seven, ranked by attention scores shown in Fig. 3C. In ISL, the top seven fingers achieved an average accuracy of 88.9%, while the bottom seven reached 77.6%. In ASL, the corresponding accuracies were 88.4 and 78.4%. The results indicate that finger-level relevance is aligned with translation performance and support the use of attention-based finger selection. Detailed per-fold results are provided in fig. S12.

Figure 3F shows heatmaps generated using relevance-weighted class activation map (R-CAM) (*36*) for the words "about" (ISL) and "animal" (ASL), illustrating the temporal regions to which the model assigns the highest relevance during classification. For about, which involves only right-hand gestures, relevance is primarily concentrated on the right-hand signal channels. In contrast, animal is a bimanual gesture, and the corresponding heatmap indicates distributed attention across both hands. In both cases, the model predominantly attends to segments with high temporal variation, demonstrating that gesture transitions with high temporal variation play a critical role in classification.

Word recognition results

Figure 4A presents confusion matrices for the 100 class classification task in both ISL and ASL. The 100 class vocabulary used in the experiments is listed in fig. S13. The matrices were computed under the unseen evaluation setting and depict the distribution of model predictions across all target classes. Most classes are well separated, and these results are consistent with the quantitative findings, further demonstrating the model's ability to distinguish a large set of sign language gestures with high accuracy. To further examine scalability with respect to vocabulary size, we performed an ablation study over subsets of the ASL vocabulary (10, 20, 50, and 100 words). The resulting mean accuracies showed the expected gradual decline as the number of classes increased (98.5, 92.25, 91.70, and 88.02%, respectively) while still maintaining high performance at 100 words. The detailed results are provided in fig. S14. In addition, to assess robustness to variation in signing speed, we performed a controlled experiment comparing deliberately slow (>2 s) and fast (<2 s) executions of four representative words and observed stable recognition performance across both conditions, as shown in fig. S15. Figure 4B illustrates the t-distributed stochastic neighbor embedding (t-SNE) mapping of the learned feature representations for both ISL (left) and ASL (right), evaluated under the unseen setting. Text features are shown as × and sensor-derived features as *. For most classes, sensor features are closely distributed around the corresponding text features, indicating that the model effectively aligns multimodal representations across modalities. Furthermore, the features form well-separated clusters for each of the 100 target words, demonstrating that the learned embedding space preserves both interclass separability and cross-modal consistency. These results are further supported by movie S3, which presents real-time demonstrations of the model accurately recognizing a range of sign language gestures across multiple word classes using off-board processing environment such as personal computers that receives wireless sensor streams and performs deep learning inference in real time.

To evaluate the effectiveness of the proposed model, we compared its performance with a conventional convolutional neural network (CNN)-based baseline, as shown in Fig. 4C and fig. S16. Experiments were conducted on both ISL and ASL datasets using three unseen users out of five participants and fourfold cross-validation. In ISL, the proposed model achieved an average accuracy of 88.9%, representing an improvement of 2.3 percentage points over the 86.6% obtained by the CNN baseline. In ASL, the proposed model attained 88.4%, which is 3.0 percentage points higher than the 85.4% of the CNN baseline.

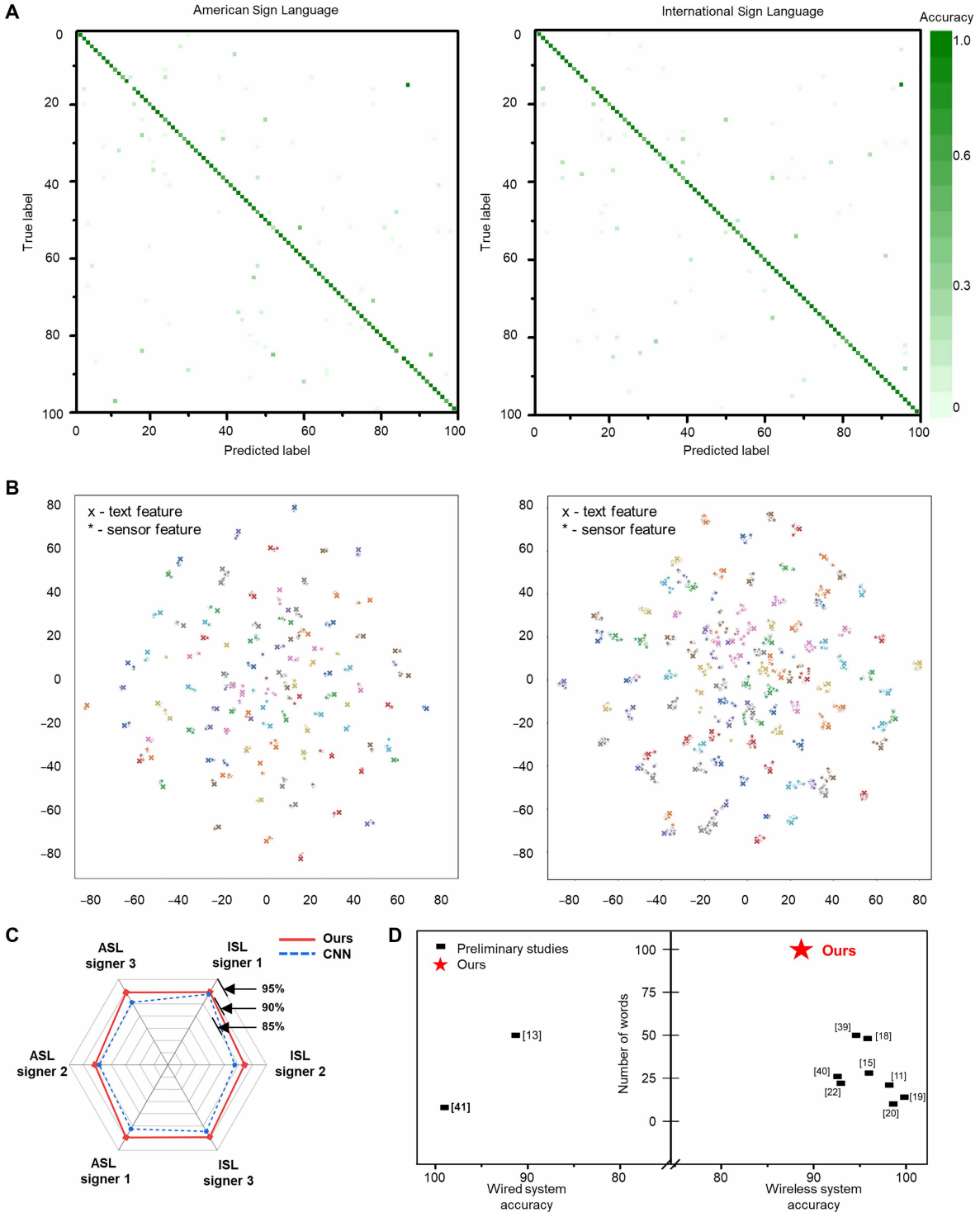

Fig. 4. Model evaluation under user-independent setting. (**A**) Confusion matrices for 100-word classification in ASL (left) and ISL (right), showing high accuracy and minimal interclass confusion under unseen-user evaluation. (**B**) t-SNE visualization of text (×) and sensor (*) embeddings for ASL (left) and ISL (right), demonstrating cross-modal alignment and class-wise separability. (**C**) Radar plot comparing classification accuracy of the proposed model and CNN baseline across three unseen ASL and ISL signers. (**D**) Benchmark plot comparing word count and recognition accuracy of prior wireless sign language systems. WRSLT achieves the highest vocabulary size (total 200 words: 100 ASL + 100 ISL) with competitive accuracy in an unseen-user setting, enabled by its wirelessly connected, ring-type, multilink architecture. References are listed next: (*11*, *13*, *15*, *18*–*20*, *22*, *39*–*41*). a.u., arbitrary unit.

These results demonstrate the improved generalization capability of the proposed architecture for user-independent sign language translation.

Figure 4D summarizes the relationship between training size and recognition accuracy reported across previously published wireless sign language translation systems. Most existing approaches have been limited to trained vocabularies of fewer than 50 words despite achieving relatively high classification accuracy. In contrast, the proposed WRSLT substantially expands the vocabulary size, demonstrating robust performance on a 200 word dataset comprising 100 words each from ASL and ISL. This scale represents more than a twofold increase in class diversity compared to prior wireless systems. The WRSLT platform maintain high recognition accuracy under an unseen evaluation setting, in which only two signers were used for training and all test participants were excluded from the training set. Despite the increased vocabulary size and limited training data, the system achieved competitive accuracy, indicating strong generalizability and user independence. In addition to performance scalability, WRSLT is the first to integrate a fully wireless, ring-type architecture with inter-ring connectivity based on multilink Bluetooth communication. Unlike systems that rely on centralized modules or partial wiring between sensors, WRSLT preserves full finger mobility and dynamic gesture expression without mechanical constraint. These results collectively highlight the scalability, robustness, and practicality of the WRSLT for real-world, large-scale sign language translation.

Sentence-level recognition using sequential word detection framework

Sign language communication occurs in the form of complete sentences, which are composed of continuous sequences of gestures. To support sentence-level translation, one approach is to train models on full gesture sequences corresponding to specific sentences (nonsegmentation-based learning). However, this method is fundamentally limited by scalability, as it would require exhaustive training across an impractically large number of possible sentence combinations. To overcome this, we implemented a segmentation-based approach in which each word in the sentence is recognized based on the trained words. Sentence-level translation is then achieved by continuously detecting words from a stream of signing input using a sequential word detection framework developed in this study. This framework builds upon a window-sliding mechanism, allowing the system to identify and segment individual words from continuous gesture sequences in real time, without requiring prior knowledge of sentence boundaries.

To enable real-time sentence construction from continuous signing, a window-sliding technique was used as part of the proposed sequential word detection framework. The sliding windows were defined with a fixed width of 200 frames and a step size of 25 frames between consecutive windows. This process was applied simultaneously across all seven selected fingers. Figure 5A presents a representative example of this process using the R1 sensor data from the sentence "Boy forget friend name." In this framework, a word is recognized as an intended gesture only when it is detected in two or more consecutive sliding windows. For instance, in the early segment of the input sequence, the word "boy" is consistently identified at frame indices 50, 75, and 100. When the output at frame 125 deviates from this pattern, "boy" is finalized as the intended word of the signer. Similarly, "forget" is identified in three consecutive windows, "friend" in four, and "name" in two, confirming their intended status based on consistent repetition. Figure 5B presents a magnified view of the forget segment from the full sentence-level sliding window detection illustrated in Fig. 5A. Once the word boy is finalized on the basis of its last repeated occurrence, a new series of sliding windows is initiated from that point—frame index 300 in this example. The subsequent windows (starting at frames 300, 325, and 350) initially yield sporadic predictions such as "start," "family," and "know," which are classified as candidate words due to their inconsistent appearance. These segments typically correspond to transitional hand movements between distinct sign gestures, which are not explicitly represented in the training dataset. From the fourth window onward (frame 375), the trained word forget begins to appear consistently across three consecutive windows, satisfying the criteria for repetition-based confirmation. At frame 450, a new prediction "rest" appears, indicating a transition out of the gesture sequence for forget. As a result, forget is accepted as the intended word, and the system identifies this boundary to trigger the next round of window sliding from the end of the last confirmed segment.

To evaluate sentence-level performance, we applied the proposed sequential word detection framework to five representative sentences (S1 to S5) constructed from 17 distinct ASL words selected from a total of 100 trained vocabulary items. These sentences were constructed to form semantically valid expressions from the 100 pretrained vocabulary set, and the real-time recognition results are demonstrated in movie S4. In our central processing unit (CPU)-based implementation, this configuration results in an average network inference time of ~0.10 s for single-word translation and about 0.36 s for the sentence-level examples while requiring only 14.3 mega floating-point operations per second (MFLOPS) per inference. As shown in Fig. 5C, each sentence was encoded into a continuous hand gesture sequence and processed through sliding window segmentation. All intended words in the sentences appeared with two or more consecutive detections, satisfying the criterion for valid word recognition. Nontarget intervals generated only isolated or inconsistent outputs, which were effectively rejected by the repetition-based decision rule. The corresponding confusion matrix for the 17 selected words used in sentences is presented in Fig. 5D, illustrating high classification accuracy and minimal confusion between words. Nonrelevant or transitional gestures were excluded from the confusion matrix. Furthermore, as demonstrated in fig. S17, the model maintained robust performance even when the order of words in the sentences was shuffled, confirming its generalizability to various sentence compositions based on continuous sequences of recognized words. Figure 5E shows word-level accuracy across the five test sentences, with all exceeding 90%, indicating reliable performance under continuous signing conditions. Figure 5F summarizes sentence-level accuracy across multiple repetitions, highlighting the consistency with which the system reconstructed the intended sequence of words. These results demonstrate that the proposed sequential word detection framework not only enables robust segmentation of gesture streams but also accurately reconstructs full sentences by leveraging a repetition-driven word validation strategy. The approach supports real-time sign language translation with strong generalizability and no reliance on predefined sentence structures, making it suitable for practical and scalable deployment.

DISCUSSION

Prior sign language translation systems have provided a valuable foundation for gesture recognition by enabling the integration of

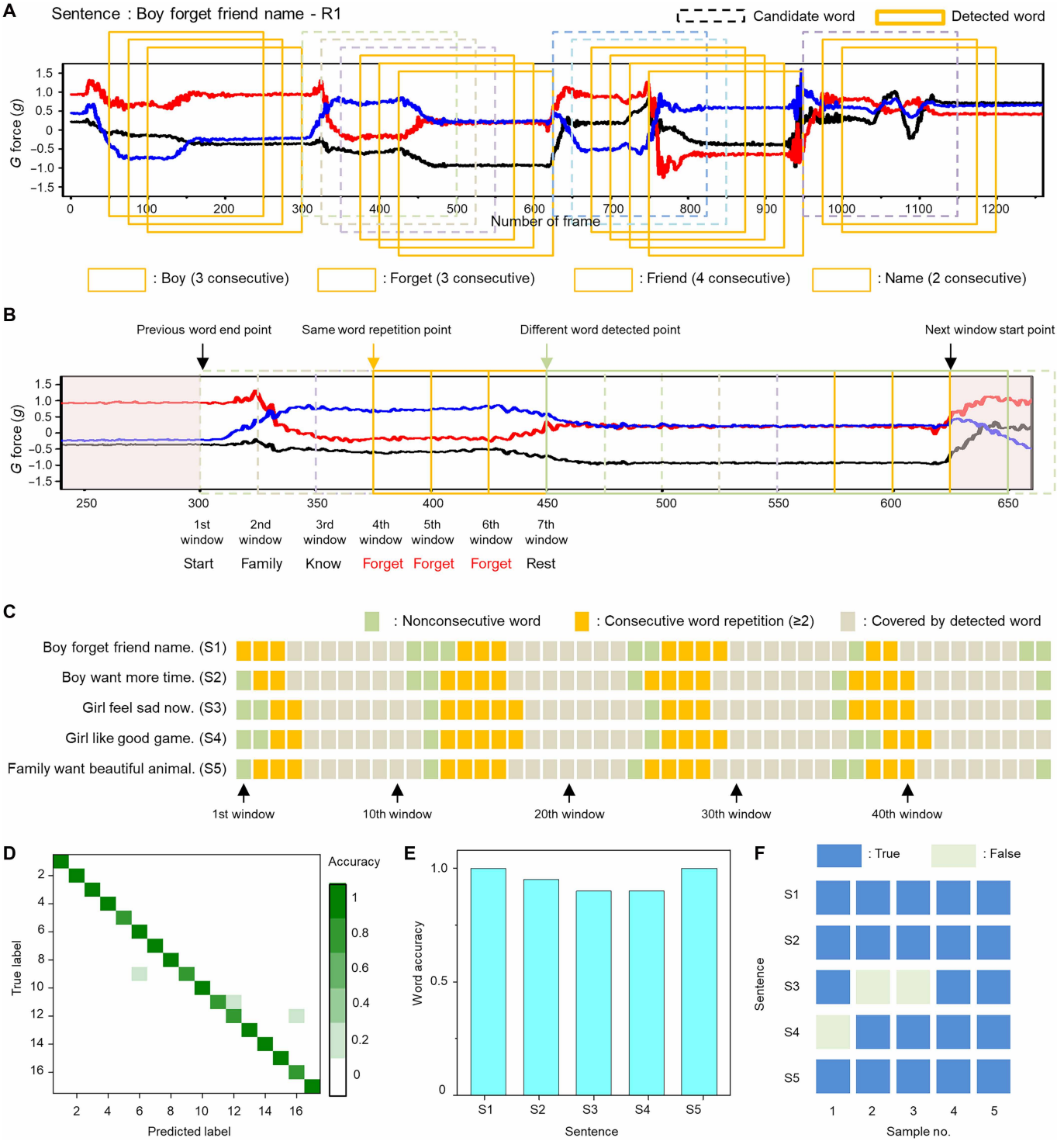

Fig. 5. Sentence-level sign language recognition using sequential word detection. (**A**) Sliding window-based detection framework applied to the sentence "Boy forget friend name" using R1 sensor data. Words are finalized only when detected in two or more consecutive windows, enabling robust segmentation of continuous gesture input. (**B**) Enlarged view of the forget segment, illustrating how repeated detection stabilizes word recognition and how transitions between gestures are handled as candidate (nonfinalized) outputs. (**C**) Sequential word detection framework detection results for five ASL sentences (S1 to S5), with valid words identified through repeated occurrences (≥2). (**D**) Confusion matrix for 17 target words used in sentence construction, showing high classification accuracy. A single instance of misclassification as "talk"—a word not included in the sentence—was observed but not reflected in the matrix. (**E**) Word-level recognition accuracy across test sentences, all above 90%. (**F**) Sentence-level recognition across five trials, confirming robust reconstruction of intended word sequences.

multiple sensors across the hand. However, their bulky structure, poor breathability, and lack of adaptability to varying hand anatomies limit wearability, user comfort, and long-term usability. In response to these limitations, we developed a compact, ring-type system that enables independent sensor placement on each finger while minimizing mechanical constraint. Furthermore, unlike many existing platforms that still rely on centralized modules and physical interconnections between sensors, the proposed WRSLT uses a fully wireless architecture using Bluetooth multilink communication, enabling true inter-ring connectivity and unrestricted hand motion. In particular, the use of inertial sensors on each ring allows the system to detect both static postures by measuring gravitational orientation and dynamic gestures through acceleration changes without relying on visual line-of-sight or external references. This sensing capability is essential for capturing the full complexity of sign language, which often involves fluid transitions between static and dynamic components. As a result, the WRSLT maintains robust recognition accuracy in user-independent unseen settings, as demonstrated by its performance with test subjects who were not included in the training dataset. Although the current model was trained using data from only two participants, the observed results suggest that the system's accuracy can be further improved with larger-scale training datasets. As the model continues to generalize with additional data, it is expected that accurate translation will be achievable with even fewer sensor nodes, enabling further miniaturization without sacrificing performance. Moreover, our experiments on both ASL and ISl demonstrate that the proposed framework generalizes well across different lexical sets and sign conventions, suggesting its extensibility to other sign language systems with minimal architectural modification. This multilingual adaptability highlights the potential of WRSLT as a scalable platform for diverse communication contexts worldwide.

The integration of WRSLT with deep learning highlights the importance of a multidisciplinary approach that combines wearable bioelectronics, inertial sensing, and artificial intelligence. In addition to accurate word-level classification, the system also enables sentence-level translation through a sequential detection framework, paving the way for real-time, fluid sign communication. These advances suggest that WRSLT can serve as a foundational platform for future human-machine interface (HMI) systems capable of facilitating seamless interaction between signers and nonsigners across a broad range of practical contexts, specifically enabling barrier-free public translation systems for unseen users and unrestricted daily assistive interfaces. Beyond sign language translation, the ring-type, wireless, and modular architecture of WRSLT may also be extended to other gesture-driven applications such as virtual or augmented reality control, touchless device interfaces, or rehabilitation monitoring systems where fine-grained hand movement tracking is essential.

MATERIALS AND METHODS

Device fabrication

The fabrication process is described in fig. S18. The fabrication of the ring-type sign language translation device begins with a Cu/PI/Cu (12.5 μm/25 μm/12.5 μm) trilayer film, from which one copper layer is completely etched away. The resulting asymmetric structure is then laminated onto a degassed polydimethylsiloxane (PDMS) (10:1 base to curing agent ratio) layer precoated on a clean glass substrate. Laser cutting is performed using an LPKF U4 Laser system to define the detailed device geometry. Following the laser structuring process, the patterned structure is peeled off using a thermal release tape. The detached device is subsequently transferred onto a flat PDMS substrate using an oxide bonding process to ensure mechanical adhesion (*37*). Key electronic components, including the Bluetooth SoC (MBN52832) and voltage regulator, are soldered onto the designated pads. Two neodymium magnets are bonded to the PDMS substrate using adhesive glue. To encapsulate the assembled device, a PDMS dip-coating process is applied, forming a flexible and biocompatible protective layer. Last, the device is cut into a ring configuration and mechanically shaped to conform to a cylindrical form factor suitable for finger attachment.

Software code uploading and device performance test

The firmware for the Bluetooth SoC (MBN52832) was programmed using Segger Embedded Studio. For real-time smartphone communication, a custom Android application was developed using Kotlin, allowing wireless control and data visualization (movie S1). In addition, for PC-based demonstrations, the aggregated data from multiple WRSLT devices were transmitted to an external computer acting as a central node, where a Python-based script was used to process and visualize the recognition results in real time (movie S2). To quantitatively evaluate WRSLT's response to varying angular positions and motion speeds, a programmable motorized setup (Arduino Uno and MG995 servo motor) was used. The motor motion was controlled via an Arduino-based code, and the WRSLT was mounted on the motor to systematically collect gesture-related inertial data under predefined movement conditions.

Algorithms for sign language translation

The architecture of the sign language translation model is based on a multimodal framework that combines a pretrained language encoder and a custom-designed signal encoder. The text encoder, built on a transformer-based foundation model, generates semantic embeddings $t_i \in R^d$ for each target word i. To improve semantic stability and reduce prompt sensitivity, we utilize an ensemble approach using multiple natural language prompts per word. Specifically, for each word, four distinct textual formulations (e.g., "A gesture that represents the concept of 'word,'" "The meaning of this sign is 'word'" were used as input to the pretrained language model. The resulting embeddings were averaged to form a single ensemble representation for each class prototype

$$t_i = \frac{1}{N}\sum_{j=1}^{N} f_{\text{text}}\left(P_j^i\right)$$

where P_j^i denotes the j-th prompt for word i and $f_{\text{text}}(\cdot)$ is the language encoder. To mitigate overfitting and promote alignment flexibility, isotropic Gaussian noise is injected into the text features during training, producing perturbed embeddings $\tilde{t}_i = t_i + \epsilon$, where $\epsilon \sim \mathcal{N}(0, I)$. The signal encoder takes time-series inertial data x_i from seven finger-mounted sensors. The raw inputs are first quantized into five discrete levels based on predefined thresholds. The quantized data then pass through a linear layer to produce features. These features are concatenated with features extracted by convolutional layers to yield the final sensor embedding $s_i \in R^d$. This embedding is projected into a shared latent space with text embeddings through fully connected layers. A categorical cross-entropy loss is applied for gesture classification over the vocabulary of K words

$$\mathcal{L}_{\text{cls}} = -\sum_{i=1}^{K} y_i \log \hat{y}_i$$

where y_i is the ground-truth label and $\hat{y}_i$ is the predicted softmax output from the classifier. To align the gesture features with the semantic space, we propose a cosine similarity loss between sensor and perturbed text embeddings

$$\mathcal{L}_{\text{c}} = 1 - \cos\left(s_i, \tilde{t}_i\right) = 1 - \frac{s_i \cdot \tilde{t}_i}{|s_i||\tilde{t}_i|}$$

The total loss is a sum of the two components

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cls}} + \mathcal{L}_{\text{c}}$$

This dual-loss formulation enables the model to learn both discriminative features for accurate classification and semantically grounded representations for robust generalization. During inference, only the signal encoder and classifier are used, allowing for real-time word prediction from sensor data alone, without any reliance on text inputs.

Data acquisition and demonstration process

To get a training dataset of the gravity-based sign language translation system, we constructed a wired module using Arduino-based system configured identically to the components integrated into WRSLT (fig. S19). The data were initially collected from all 10 fingers under quasi-static conditions of the whole body, followed by a post hoc ablation study to evaluate finger-wise importance. The sign language vocabulary was curated based on commonly used signs presented on the HandSpeak platform (www.handspeak.com) and selected to ensure availability of corresponding signs in both ASL and ISL as referenced on the sign language dictionary platform (https://spreadthesign.com). Two participants were selected for training data collection. Each participant performed 50 repetitions for every word in the vocabulary, resulting in a total of 100 training samples per word. For user-independent testing, five additional participants—distinct from the training group—were recruited. This setup, in which a large number of unseen participants are used for testing than for training, ensures rigorous user-independent evaluation. Each test participant wore the WRSLT and repeated each word five times, yielding 15 test samples per word for evaluating generalization to unseen users. Sentence-level gesture data were collected from the train participants to assess sequential recognition performance. Real-time demonstrations (movies S3 and S4) were performed by streaming all seven wireless sensor channels to a single external host PC, where the proposed model executed inference in real time to produce interactive outputs.

Mechanical simulation

Strain distributions for both serpentine and straight interconnect structures are illustrated in figs. S3 and S5. Finite element simulations were conducted using Ansys Mechanical 2021 R2 to analyze the mechanical behavior of both straight and serpentine interconnect structures based on material properties and layered configurations. The elastic modulus and Poisson's ratio were set to 2 GPa and 0.34 for 25-μm-thick polyimide, and 126 GPa and 0.345 for 12.5-μm-thick Cu, respectively. The right edge of the structure was constrained with a fixed support, while a displacement was applied to the opposite edge, causing the structure to bend toward the center. The equivalent elastic strain was evaluated over a range of 0 to 2%. This result showed that the straight interconnects experienced strains close to the yield limit of copper (~3%) (*38*) under bending. However, the serpentine structures accommodated the deformation more effectively, enabling mechanically stable bending behavior.

Supplementary Materials

The PDF file includes:
Figs. S1 to S19
Table S1
Legends for movies S1 to S4

Other Supplementary Material for this manuscript includes the following:
Movies S1 to S4

REFERENCES

1. M. M. Nomeland, R. E. Nomeland, *The deaf community in America: History in the making* (McFarland, 2011).
2. J. Fenlon, E. Wilkinson, "Sign languages in the world," in *Sociolinguistics and Deaf Communities*, A. C. Schembri, C. Lucas, Eds. (Cambridge Univ. Press, 2015), pp. 5–28.
3. K. Kudrinko, E. Flavin, X. Zhu, Q. Li, Wearable sensor-based sign language recognition: A comprehensive review. *IEEE Rev. Biomed. Eng.* **14**, 82–97 (2021).
4. S. Sharma, S. Singh, Vision-based hand gesture recognition using deep learning for the interpretation of sign language. *Expert Syst. Appl.* **182**, 115657 (2021).
5. D. Sarma, M. K. Bhuyan, Methods, databases and recent advancement of vision-based hand gesture recognition for hci systems: A review. *SN Comput. Sci.* **2**, 436 (2021).
6. M. M. Zaki, S. I. Shaheen, Sign language recognition using a combination of new vision based features. *Pattern Recogn. Lett.* **32**, 572–577 (2011).
7. J. Wu, L. Sun, R. Jafari, A wearable system for recognizing American sign language in real-time using IMU and surface EMG sensors. *IEEE J. Biomed. Health Inform.* **20**, 1281–1290 (2016).
8. J. Wu, Z. Tian, L. Sun, L. Estevez, R. Jafari, "Real-time American Sign Language recognition using wrist-worn motion and surface EMG sensors," in *2015 IEEE 12th international conference on wearable and implantable body sensor networks (BSN)* (IEEE, 2015), pp. 1–6.
9. Q. Zhang, J. Jing, D. Wang, R. Zhao, Wearsign: Pushing the limit of sign language translation using inertial and emg wearables. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **6**, 1–27 (2022).
10. B. G. Lee, S. M. Lee, Smart wearable hand device for sign language interpretation system with sensors fusion. *IEEE Sensors J.* **18**, 1224–1232 (2017).
11. X. Wu, X. Luo, Z. Song, Y. Bai, B. Zhang, G. Zhang, Ultra-robust and sensitive flexible strain sensor for real-time and wearable sign language translation. *Adv. Funct. Mater.* **33**, 2303504 (2023).
12. S. Lee, D. Jo, K.-B. Kim, J. Jang, W. Park, Wearable sign language translation system using strain sensors. *Sens. Actuators A* **331**, 113010 (2021).
13. F. Wen, Z. Zhang, T. He, C. Lee, AI enabled sign language recognition and VR space bidirectional communication using triboelectric smart glove. *Nat. Commun.* **12**, 5378 (2021).
14. P. Maharjan, T. Bhatta, J. Y. Park, "Self-powered hybridized kinetic energy harvester for human motion based on frequency up-conversion mechanism," in *2019 20th International Conference on Solid-State Sensors, Actuators and Microsystems & Eurosensors XXXIII (TRANSDUCERS & EUROSENSORS XXXIII)* (IEEE, 2019), pp. 385–388.
15. A. Qaroush, S. Yassin, A. Al-Nubani, A. Alqam, Smart, comfortable wearable system for recognizing Arabic Sign Language in real-time using IMUs and features-based fusion. *Expert Syst. Appl.* **184**, 115448 (2021).
16. M. Ahmed, B. Zaidan, A. Zaidan, M. M. Salih, Z. Al-Qaysi, A. Alamoodi, Based on wearable sensory device in 3D-printed humanoid: A new real-time sign language recognition system. *Measurement* **168**, 108431 (2021).
17. C. Yiu, Y. Liu, W. Park, J. Li, X. Huang, K. Yao, Y. Gao, G. Zhao, H. Chu, J. Zhou, Skin-interfaced multimodal sensing and tactile feedback system as enhanced human-machine interface for closed-loop drone control. *Sci. Adv.* **11**, eadt6041 (2025).
18. Y. Liu, X. Jiang, X. Yu, H. Ye, C. Ma, W. Wang, Y. Hu, A wearable system for sign language recognition enabled by a convolutional neural network. *Nano Energy* **116**, 108767 (2023).
19. Z. Sun, M. Zhu, X. Shan, C. Lee, Augmented tactile-perception and haptic-feedback rings as human-machine interfaces aiming for immersive interactions. *Nat. Commun.* **13**, 5224 (2022).
20. Z. Zhou, K. Chen, X. Li, S. Zhang, Y. Wu, Y. Zhou, K. Meng, C. Sun, Q. He, W. Fan, Sign-to-speech translation using machine-learning-assisted stretchable sensor arrays. *Nat. Electron.* **3**, 571–578 (2020).

21. D. Yao, W. Wang, H. Wang, Y. Luo, H. Ding, Y. Gu, H. Wu, K. Tao, B.-R. Yang, S. Pan, J. Fu, F. Huo, J. Wu, Ultrasensitive and breathable hydrogel fiber-based strain sensors enabled by customized crack design for wireless sign language recognition. *Adv. Funct. Mater.* **35**, 2416482 (2025).
22. C. K. Mummadi, F. Philips Peter Leo, K. Deep Verma, S. Kasireddy, P. M. Scholl, J. Kempfle, K. Van Laerhoven, "Real-time and embedded detection of hand gestures with an IMU-based glove," in *Informatics*. (MDPI, 2018), vol. 5, pp. 28.
23. N. M. Kakoty, M. D. Sharma, Recognition of sign language alphabets and numbers based on hand kinematics using a data glove. *Procedia Comput. Sci.* **133**, 55–62 (2018).
24. F. Pezzuoli, D. Corona, M. L. Corradini, Recognition and classification of dynamic hand gestures by a wearable data-glove. *SN Comput. Sci.* **2**, 5 (2021).
25. N. Praveen, N. Karanth, M. Megha, "Sign language interpreter using a smart glove," in *2014 international conference on advances in electronics computers and communications* (IEEE, 2014), pp. 1–5.
26. A. Z. Shukor, M. F. Miskon, M. H. Jamaluddin, F. bin Ali, M. F. Asyraf, M. B. bin Bahar, A new data glove approach for Malaysian sign language detection. *Procedia Comput. Sci.* **76**, 60–67 (2015).
27. K. S. Abhishek, L. C. F. Qubeley, D. Ho, "Glove-based hand gesture recognition sign language translator using Arduino," in *2016 IEEE international conference on electron devices and solid-state circuits (EDSSC)* (IEEE, 2016), pp. 334–337.
28. L. Li, S. Jiang, P. B. Shull, G. Gu, SkinGest: artificial skin for gesture recognition via filmy stretchable strain sensors. *Adv. Robot.* **32**, 1112–1121 (2018).
29. C. Yan, S. Jiang, Y. Wang, J. Deng, X. Wang, Z. Chen, T. Chen, H. Huang, H. Wu, A wearable sign language translation device utilizing silicone-hydrogel hybrid triboelectric sensor arrays and machine learning. *Nano Energy* **133**, 110425 (2025).
30. H.-J. Kim, S.-W. Baek, Application of wearable gloves for assisted learning of sign language using artificial neural networks. *Processes* **11**, 1065 (2023).
31. R. Wu, S. Seo, L. Ma, J. Bae, T. Kim, Full-fiber auxetic-interlaced yarn sensor for sign-language translation glove assisted by artificial neural network. *Nano-Micro Lett.* **14**, 139 (2022).
32. M. A. A. Faisal, F. F. Abir, M. U. Ahmed, M. A. R. Ahad, Exploiting domain transformation and deep learning for hand gesture recognition using a low-cost dataglove. *Sci. Rep.* **12**, 21446 (2022).
33. K. R. Pyun, K. Kwon, M. J. Yoo, K. K. Kim, D. Gong, W.-H. Yeo, S. Han, S. H. Ko, Machine-learned wearable sensors for real-time hand-motion recognition: toward practical applications. *Natl. Sci. Rev.* **11**, nwad298 (2024).
34. A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, "Learning transferable visual models from natural language supervision," in *International conference on machine learning* (PmLR, 2021), pp. 8748–8763.
35. A. Binder, G. Montavon, S. Lapuschkin, K.-R. Müller, W. Samek, "Layer-wise relevance propagation for neural networks with local renormalization layers," in *Artificial Neural Networks and Machine Learning–ICANN 2016: 25th International Conference on Artificial Neural Networks, Part II 25* (Springer, 2016), pp. 63–71.
36. J. R. Lee, S. Kim, I. Park, T. Eo, D. Hwang, "Relevance-cam: Your model already knows where to look," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 14944–14953.
37. J. Park, H. W. Kim, S. Lim, H. Yi, Z. Wu, I. G. Kwon, W. H. Yeo, E. Song, K. J. Yu, Conformal fixation strategies and bioadhesives for soft bioelectronics. *Adv. Funct. Mater.* **34**, 2313728 (2024).
38. J. Park, K. Kim, Y. Kim, T. S. Kim, I. S. Min, B. Li, Y. U. Cho, C. Lee, J. Y. Lee, Y. Gao, A wireless, solar-powered, optoelectronic system for spatial restriction-free long-term optogenetic neuromodulations. *Sci. Adv.* **9**, eadi8918 (2023).
39. A. Tashakori, Z. Jiang, A. Servati, S. Soltanian, H. Narayana, K. Le, C. Nakayama, C.-l. Yang, Z. J. Wang, J. J. Eng, Capturing complex hand movements and object interactions using machine learning-powered stretchable smart textile gloves. *Nat. Mach. Intell.* **6**, 106–118 (2024).
40. P. Tan, X. Han, Y. Zou, X. Qu, J. Xue, T. Li, Y. Wang, R. Luo, X. Cui, Y. Xi, Self-powered gesture recognition wristband enabled by machine learning for full keyboard and multicommand input. *Adv. Mater.* **34**, e2200793 (2022).
41. S. Oh, J.-I. Cho, B. H. Lee, S. Seo, J.-H. Lee, H. Choo, K. Heo, S. Y. Lee, J.-H. Park, Flexible artificial Si-In-Zn-O/ion gel synapse and its application to sensory-neuromorphic system for sign language translation. *Sci. Adv.* **7**, eabg9450 (2021).

Acknowledgments

Funding: This work was supported by the National Research Foundation of Korea (RS-2024-00353768) (to K.J.Y.), the National Research Foundation of Korea (RS-2025-02217919 to K.J.Y. and D.H.), the National Research Foundation of Korea (RS-2025-02215070 to K.J.Y. and D.H.), the Artificial Intelligence Graduate School Program at Yonsei University (RS-2020-II201361 to D.H.), the National Research Foundation of Korea (RS-2026-25481951 to K.K.), KIST Institutional Program (2E33801 to D.H.), KIST Institutional Program (2E33800 to D.H.), the National Research Foundation of Korea (RS-2025-16070382 to D.H.), the Samsung Research Funding Center of Samsung Electronics (SRFC-IT1901-08 to K.J.Y.), the Hankuk University of Foreign Studies Research Fund of 2025 (to K.K., Y.P., D.K., and B.J.), and the Gyeonggi Regional Innovation System & Education Project (Gyeonggi RISE Project), supported by the Ministry of Education and Gyeonggi Province (to K.K. and B.J.). **Author contributions:** Conceptualization: J.P., Y.S., Y.L., J.B., J.K., Y.C., K.Ka., D.H., and K.J.Y. Methodology: J.P., Y.S., Y.L., J.B., J.K., Y.U.C., Y.C., K.Ka., and K.J.Y. Software: J.P., Y.S., I.S.M., Y.L., J.-H.H., S.H.P., B.J., and K.Ka. Validation: J.P., Y.S., Y.L., S.H.P., J.B., D.K., Y.C., K.Ka., and K.J.Y. Formal analysis: J.P., Y.S., Y.L., J.L., S.H.P., and K.Ka. Investigation: J.P., Y.S., I.S.M., Y.L., J.L., J.-H.H., S.P., J.B., J.K., Y.C., K.Ka., and K.J.Y. Resources: Y.S., I.S.M., Y.L., J.B., and K.J.Y. Data curation: J.P., Y.S., Y.L., S.H.P., Y.U.C., and K.Ka. Writing—original draft: J.P., Y.S., Y.L., Y.C., K.Ka., and K.J.Y. Writing—review and editing: J.P., Y.S., I.S.M., S.H.P., Y.P., J.K., Y.C., K.Ka., D.H., and K.J.Y. Visualization: J.P., Y.S., Y.L., J.B., T.K., K.Ki., D.K., Y.U.C., and K.Ka. Supervision: Y.C., K.Ka., D.H., and K.J.Y. Project administration: J.P., J.B., D.K., K.Ka., D.H., and K.J.Y. Funding acquisition: D.K., Y.P., B.J., K.Ka., D.H., and K.J.Y. **Competing interests:** The authors declare that they have no competing interests. **Data, code, and materials availability:** All data and code needed to evaluate and reproduce the results in the paper are present in the paper and/or the Supplementary Materials. This paper did not generate new materials. The codes are present in the GitHub (https://github.com/yejees/WRSLT). The codes and source data for figures are available from Dryad with the identifier (https://doi.org/10.5061/dryad.nvx0k6f5d).

Submitted 7 October 2025
Accepted 30 March 2026
Published 1 May 2026
10.1126/sciadv.aec8995

ScienceAdvances

An AI-driven, wearable, conformal ring system for real-time and user-independent sign language interpretation

Jaejin Park, Yejee Shin, In Sik Min, Yeeun Lee, Jiwon Lee, Jung-Hoon Hong, Sunho Park, Sang Hoon Park, Junhyeong Baek, Taemin Kim, Kyubeen Kim, Doyun Kim, Yunkyu Park, Jongyoo Kim, Young Uk Cho, Yeonsik Choi, Byunghwan Jeon, Kyowon Kang, Dosik Hwang, and Ki Jun Yu

Sci. Adv. **12** (18), eaec8995. DOI: 10.1126/sciadv.aec8995

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.