The In-Car Sign Language Corpus (ICSL): A Multi-Modal Resource for Constrained-Space Sign Language Recognition

Raviteja Boddu*, Guilherme Vieira Leite†, Joed Lopes da Silva*, Ângelo Benetti†, Isabela Barbieri†, Natália de Melo Afonso†, Thyago Santos‡, Helio Pedrini†, Felipe Venâncio Barbosa‡, José Mario De Martino†, Munir Georges*, Alessandro Zimmer*

*Technische Hochschule Ingolstadt (THI), Bayern, Germany
†Universidade Estadual de Campinas (UNICAMP), São Paulo, Brazil
‡Universidade de São Paulo (USP), São Paulo, Brazil
*{raviteja.boddu, joed.lopesdasilva, munir.georges, alessandro.zimmer}@thi.de
†martino@unicamp.br, †guilherme.leite@ic.unicamp.br

Abstract

This paper addresses the challenges of using sign language within shared mobility services, such as taxis, carpools, or ride-sharing platforms. The use of sign language recognition (SLR) in real-world, confined environments, specifically vehicle interiors remains largely unexplored. To motivate research in this area, we present the In-Car Sign Language (ICSL) dataset for Brazilian Sign Language (Libras), with the long-term goal of improving public transport accessibility for the Deaf and Hard-of-Hearing community. The dataset consists of: (1) high-precision laboratory motion capture (MoCap) data to establish an idealized linguistic baseline and (2) real-world multi-modal in-car recordings captured using a 2D camera and 3D Time-of-Flight sensors. The dataset provides a basis for comparative analyses between synthesized signing avatar animations and recorded real signing interpreter videos, which enable future research into robust "in-the-wild" SLR models and domain adaptation. We describe in detail the use cases, the setup, the data collection protocol, and the metadata structure of the corpus. In total, we recorded a multimodal dataset exceeding 1.5 million frames, comprising the synchronized multimodal streams described above featuring Libras users across various in-car scenarios. The corpus is provided with gloss annotation of lexical signs and non-lexical sign language elements specially designed to support the training and evaluation of deep neural networks for constrained space recognition. In-vehicle signing offers a technically significant example of a constrained, occluded, and non-frontal environment. While recognizing the diverse communication strategies already employed by the Deaf community, identifying automotive-specific limitations provides a useful stepping stone for research into enhancing in-car accessibility and passenger quality of life.

Keywords: Brazilian Sign Language (Libras), Shared Mobility Service, Motion Capture (MoCap), In-Car Communication, Constrained Signing Space, Signing Avatars, Multimodal Sensors

1. Introduction

The advancement in natural language processing and computer vision has brought us closer to seamless human-machine interaction. While the Deaf and Hard-of-Hearing (DHH) community has developed effective strategies for "on-the-go" communication, such as the use of mobile devices or visual cues, the interior of vehicles within shared mobility services represents a technically significant and under-researched environment for Sign Language Recognition (SLR).

Specifically, the car cabin serves as a critical case study for constrained, occluded, and non-frontal signing environments, which are often overlooked in traditional laboratory-based datasets. While Brazilian Sign Language (Libras) has an established linguistic foundation, there is currently no specialized resource documenting how signing is produced and processed in the highly constrained physical and visual environments of a car cabin (dos Santos et al., 2025; Lee et al., 2025).

The relevant studies and corpora that exist for Libras are predominantly recorded in laboratory-controlled settings with optimal lighting and neutral backgrounds. In such environments, signers have a full, unobstructed signing space (dos Santos et al., 2025). However, the interior of a vehicle introduces severe constraints, such as seatbelts cutting across the torso, occlusions, dashboard obstruction, and constrained space that limit the range of torso, elbow, and arm movements, and dynamic environmental lighting creates substantial visual noise. These factors represent fundamental challenges for current SLR systems, which often fail to generalize to such "in-the-wild" conditions.

Within the framework of Project UNITY, a collaborative research initiative between the UNICAMP (Brazil) and the THI (Germany), we seek to establish a specialized multimodal dataset that captures real-world constraints while providing a high-fidelity baseline. We aim to enhance the robustness of SLR models by utilizing signing avatars as a controlled reference to evaluate the environmental chal-

Proceedings of the 12th Workshop on the Representation and Processing of Sign Languages, pages 31–41
16 May 2026.

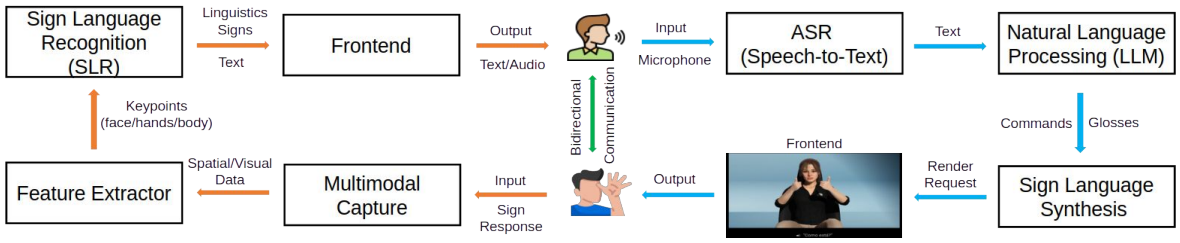


Figure 1: Proposed Project UNITY architecture for a conceptual bidirectional loop: (right) Driver-to-Passenger Speech-to-Sign Synthesis and (left) Passenger-to-Driver Sign Language Recognition.

lenges present in real-world in-car video. Led by professional Libras signers and Deaf researchers, who are also co-authors of this study, the data acquisition process seeks to document the authentic challenges of the constrained transport domain by documenting passenger-driver communication within the vehicle cabin.

The long-term vision of this project is to investigate the feasibility of a real-time, bidirectional translation framework, as seen in Figure 1, which serves as the target architecture for our ongoing work. Unlike traditional one-way approaches, our target architecture facilitates a conversational loop between driver speech and passenger signs (Libras). Building such a system requires resources that capture both idealized linguistic baselines and the severe physical restrictions of the vehicle cabin.

Rather than presenting a final recognition solution, this paper describes an initial step toward the In-Car Sign Language (ICSL) corpus, which we hope can serve as a helpful reference for future data collection efforts in the automotive domain. It details the acquisition protocol, use-case design, and setup for in-car recordings; the technical setup of the laboratory MoCap baseline; and the architecture of the signing avatar pipeline. By providing this comparative resource, we hope to contribute to future studies on accessible communication technologies for the DHH community in transportation systems.

2. Related Work

For the Brazilian Sign Language (Libras) community, there is a big gap between lab research and real-world use, especially in ride-hailing and ride-sharing services. The space, movement, and social setting within a vehicle influence how signs are produced. This section provides an overview of existing language resources, the algorithms used, and the real-world challenges involved in building a Libras dataset and tools designed specifically for use in vehicles.

2.1. The Landscape of Brazilian Sign Language (Libras) Resources

The development of language resources for Libras has changed over time. It started mainly with written documentation and later moved to video-based, multimodal data. Libras is recognized as a legal means of communication and expression in Brazil, with its own grammar and linguistic system for transmitting ideas and facts (Presidência da República, 2002).

Early work on documenting Libras mostly focused on creating dictionaries and bilingual corpora to support education and translation (de Quadros and de Sousa, 2021). A comparative summary of prominent Libras datasets and their recording environments is provided in Table 1.

2.1.1. Textual and Bilingual Corpora

One of the important recent resources for Libras is VLibras-DB (Lima et al., 2025). It is a bilingual text dataset with about 127,000 translation pairs between Brazilian Portuguese and Libras. This dataset was mainly created to help neural machine translation systems convert written text into glosses. Broadly defined, glossing is the use of written language to represent primary data, whether signed or spoken. Within SLR, it serves as a symbolic bridge, utilizing labels from a spoken language to represent the manual and non-manual components of a sign.

To build VLibras-DB, a group of 10 professional interpreters worked together to define annotation rules for representing signs using glosses that could leverage neural network training. The gloss convention established includes directional verbs, intensifiers, and negation. The dataset primarily covers manual lexical signs but does not address challenges posed by non-manual markers or timing-related information. Non-manual markers (NMMs) are linguistically meaningful articulations produced by the face, head, torso, and gaze used as part of a sign language's phonology, morphology, syntax, and prosody (Herrmann, 2014). As a result, the

VLibras-DB is not fully suitable for sign recognition tasks in real-world settings.

2.1.2. Visual and Isolated Sign Recognition Benchmarks

For isolated sign language recognition (ISLR), researchers typically use datasets such as MINDS-Libras (Rezende et al., 2021) and LIBRAS-UFOP (Cerna et al., 2021). These datasets have videos of single signs, mostly recorded in a controlled setup, with a plain background and the signer facing the camera. Recently, researchers have improved recognition performance by utilizing skeleton-based image representations paired with 2D CNN models, achieving accuracies up to 0.93 on MINDS-Libras. This technique involves transforming skeletal sequences comprising body, hand, and facial landmarks into a single spatio-temporal image, allowing a 2D CNN to recognize the sign from this unified input. This image then goes into a classifier to recognize the sign (Alves et al., 2024). However, such datasets do not work well for car environment. Most ISLR benchmarks assume the signer is facing the camera directly, which is very different from the angles you get inside a vehicle (Ranum et al., 2024). While vehicle interiors present unique spatial challenges, they share commonalities with other atypical or constrained-space signing environments. For instance, the PopSign dataset (Starner et al., 2023) explores the complexities of one-handed signing with the narrow field of view of a smartphone camera. Much like PopSign, our ICSL dataset addresses the problem of reduced signing volume, where the signer must adapt the scale of their gestures due to spatial boundaries. Our work extends this by focusing on the specific two-handed constraints, such as seatbelt occlusions and non-frontal camera perspectives, unique to the automotive cabin that are not present in handheld mobile datasets.

2.1.3. Regional and Documentary Inventories

The LIBRAS-UFSC corpus and the Libras National Inventory are focused on preserving Brazil's Deaf heritage and the language's diversity. They ensure the corpus includes people of different ages, genders, and regions, using careful methods to build it. The LIBRAS-UFSC corpus provides phonetic and phonological details, inspired by the Dutch Signbank model, and classifies signs by hand shape, movement, and location. These resources are relevant for linguistic research, but they are usually recorded in controlled settings, so they do not capture challenges such as visual noise, occlusions, or restricted movement that you would see in a car (de Quadros and de Sousa, 2021) (Lobo-Neto and Pedrini, 2024).

2.2. Sign Language Recognition Architectures and Methodologies

The technological evolution of SLR has transitioned from traditional manual feature extraction to end-to-end deep learning models. This shift is critical for handling the high variability of sign language, which is based on precise hand movements, facial expressions, and body language.

2.2.1. Feature Extraction and Temporal Modeling

Feature extraction focuses on capturing spatial information about hand shape, position, and orientation, as well as facial landmarks. Temporal modeling addresses the multichannel sign language characteristics (hands, facial expression, gaze, torso shift, head movement) presented in continuous stream of video frames (Camgoz et al., 2020; Rastgoo, 2021; Alyami et al., 2024; Taher and Zeebaree, 2025). Common architectures include:

- **CNN-BiLSTM:** Convolutional Neural Networks (CNNs) extract spatial features from individual frames, while Bidirectional Long Short-Term Memory (BiLSTM) networks learn the temporal dependencies over time. This combination is particularly effective for dynamic gestures, where the order of movements is crucial to the meaning (Lu et al., 2023).
- **Graph Convolutional Networks (GCNs):** These consider body landmarks as nodes in a graph, allowing the model to learn the structural interconnections between different joints and body parts (de Amorim et al., 2019).
- **Transformers:** Utilizing spatial and temporal attention mechanisms, Transformers can capture long-range dependencies across frames and correlations between different visual channels (Ma et al., 2024).

2.2.2. Landmark Detection and Efficiency

In real-time automotive applications, computational efficiency is as vital as accuracy. MediaPipe and OpenPose are the two most frequently cited tools for skeletal feature extraction due to their extensive joint coverage. OpenPose is often prone to temporal jitter and high latency, particularly in occluded environments (Alves et al., 2024). In contrast, MediaPipe offers a lightweight alternative that can achieve 5x faster recognition by utilizing an optimized subset of body landmarks (dos Santos et al., 2025). Research has shown that a well-chosen landmark subset (e.g., focusing on the upper body and hands) can maintain state-of-the-art accuracy while significantly reducing inference time. While efficient, MediaPipe's performance can degrade

Table 1: Overview of existing and proposed Brazilian Sign Language (Libras) datasets.

Dataset Name	Modality	Primary Task	Language	Scope	Annotations
VLibrasBD	Text/Gloss	Neural Machine Translation	Libras/BP	127k Pairs	Gloss, Syntax
MINDS-Libras	RGB Video	Isolated Sign Recognition	Libras	1,200+ Samples	Gloss
LIBRAS-UFOP	RGB Video	Isolated Sign Recognition	Libras	2,000+ Samples	Gloss
LIBRAS-UFSC	Video/Bank	Linguistic Documentation	Libras	National Inventory	Phonetics
LSWH100	Synthetic RGB	Handshape Classification	Libras	144k Images	Keypoints
Ours (ICSL)	**RGB/Depth/IR /Point Cloud**	**Constrained Sign Recognition**	**Libras**	**1.5M+ Frames, 127 Phrases***	**Gloss**

*Current subset of a total 1,344 defined phrases.

significantly when parts of the body are blocked or in poorly lit environments, which often happens inside a car (Amalfitano et al., 2023). Consequently, there is a need for models that can handle missing hand or body points or combine information from different sensors to deal with visual noise.

2.3. The Automotive Domain: Physical and Social Constraints

The interior of a vehicle presents a highly restricted physical and visual environment for communication. Unlike laboratory settings, where signers have full, unobstructed signing space, the car cabin introduces several "in-the-wild" constraints that serve as a starting point for specialized SLR research.

2.3.1. Constrained Signing Space and Occlusions

The physical setup inside a car changes how people sign. Seatbelts cross the body and can obscure important parts of a sign and hinder torso movements. The dashboard, headrests, and tight space also limit the movement of the arms and elbows, reducing overall signing space.

2.3.2. Environmental Noise and Dynamic Lighting

Inside a vehicle, visual conditions can change significantly. Lighting can go from bright sunlight to near darkness very quickly. Shadows, reflections from windows, and uneven light can add visual noise, making it harder to clearly see hands and movements. Because of this, sign language recognition systems that work well in lab lighting often struggle to handle these changing real-world conditions (Amalfitano et al., 2023).

2.4. Hardware and Multimodal Sensing

Because single RGB cameras do not work reliably inside a car, researchers are using multiple sensor types to make sign language recognition systems more stable and accurate.

2.4.1. 2D vs. 3D Sensing

While 2D cameras are affordable and widely available in modern vehicles, they lack depth information. Time-of-Flight (TOF)/3D cameras provide 3D skeletal data that can be used to track hand motion trajectories in three dimensions. Depth maps enable the system to segment hands from the background more effectively, especially when the hands are similar in color to the vehicle interior. Depth sensing is critical for sign language because the movement of the hand toward or away from the camera often distinguishes between different signs (Zhang, 2022) (Ranum et al., 2024).

2.4.2. Synthetic Data and Motion Capture (MoCap)

Synthetic data generation has emerged as a key strategy for overcoming the difficulty and high cost of collecting large-scale natural datasets.

Previous studies have successfully utilized motion capture systems to acquire precise three-dimensional point cloud and skeletal data from deaf signers (Kipp et al., 2011; De Martino et al., 2016; Will et al., 2018; Jedlička et al., 2020; Andersen et al., 2025; De Martino et al., 2025). This high-fidelity spatiotemporal data, capturing both manual and non-manual signals, serves as a critical training corpus for neural networks, such as CNNs and RNNs, enabling them to learn the mapping from continuous sign language kinematics to written language translation.

Shterionov et al. (2024) discuss about the types of synthetic work, two of the main trends in synthesis are: (i) the more traditional 3D animation-based which resolves in generating a 3D animated character, commonly referred to as an avatar; and (ii) a video of a virtual human that can be synthesized with generative AI methods based on real human video/image data

3. Motivation for the Set-up

The synthesis of the existing literature reveals the primary motivations for developing our specialized Libras resource. This work distinguishes itself by starting from a stable, fundamental baseline, which is currently missing in the field of automotive sign language recognition.

- **Establishing a Multimodal In-Car Baseline:** There is currently no resource documenting Libras production within the physical and visual constraints of a car cabin. Our setup provides the very first multimodal baseline (RGB + 3D TOF) recorded inside a vehicle. By starting with a simplified, stable environment, we establish the necessary scientific foundation before introducing more complex parameters in future iterations, and it provides the first comparative benchmark between idealized laboratory MoCap and these constrained "in-the-wild" conditions.
- **Comparative Analysis of Constrained Space:** This resource is motivated by the need to quantify the linguistic and physical differences between idealized laboratory signing and signing within the confined space of a car cabin. This dataset allows for a direct comparison between high-precision MoCap and real-world cabin constraints.
- **Facilitating Initial Domain Adaptation:** By providing a synchronized real-world baseline and a corresponding synthetic dataset, we enable researchers to begin exploring domain adaptation for cars using expressive 3D avatars.

3.1. Potential Research Directions

To address these motivations, the ICSL dataset provides the necessary multimodal data to investigate several key research questions in future studies, such as the following:

- **Camera Placement Optimization:** Determining the optimal camera positions within a vehicle cabin to maximize the visibility of facial expressions and head movements in Libras, despite the spatial constraints.
- **Physical Baseline Comparison:** Analyzing how sign production (hand configuration and path movement) differs when comparing a high-precision laboratory MoCap baseline to the constrained physical space of a vehicle.
- **Multimodal Stability Evaluation:** Assess to what extent multimodal sensor fusion improves the stability of landmark detection in a confined interior compared to traditional monocular systems.
- **Perspective Gap Assessment:** Evaluating whether signing avatars, generated from idealized MoCap, can effectively simulate the non-frontal viewing angles required for in-car Human-Machine Interfaces (HMI).
- **Linguistic Adaptation Analysis:** Investigating how physical restrictions, such as the car seat and seatbelt, impact the duration and spatial volume of signs compared to laboratory-recorded data.
- **Model Robustness Testing:** Quantifying the performance drop when models trained on stable in-car data are tested against other driving-related micro-gestures or environmental noise.

By establishing this baseline, we provide the scientific community with a foundation for building inclusive, real-time assistive technologies that bridge the communication gap in shared mobility services.

4. Corpus Design

The design of the In-Car Sign Language (ICSL) corpus is driven by the need for accessible passenger-driver communication within shared mobility services. Unlike general-purpose sign language datasets, this resource focuses exclusively on the constrained signing space of vehicle interiors to analyze how physical constraints impact LIBRAS language.

4.1. Scenario Selection and Use Cases

In collaboration with our Deaf researchers and professional Libras interpreters, we defined 1,344 essential communication scenarios for shared mobility services. To date, we have recorded 127 essential cases that reflect the immediate needs of DHH passengers. These are categorized into five functional groups:

- **Navigation and Routing**
  - "Qual o destino?" (Where to?)
  - "Melhor evitar o centro." (It is best to avoid downtown.)
  - "Já passamos o ponto." (We are past my drop-off.)

  - etc.

- **Safety and Regulations**
  - “Reduzindo por segurança.” (Reducing for safety reasons.)
  - “Pode ir devagar?” (Can you slow down?)
  - etc.

- **Passenger Comfort and Interaction**
  - “Preciso descer urgente.” (I need to stop urgently.)
  - “Posso ligar o ar-condicionado?” (May I turn on the air conditioning?)
  - etc.

- **Assistance**
  - “Pode ajudar com a cadeira de rodas?” (Can you help with the wheelchair?)
  - “Tem troco para R$ 25,00?” (Do you have change for R$ 25,00?)
  - “Tenho uma mala pesada.” (My luggage is heavy.)
  - etc.

- **Alphabet Spelling**
  - A, B, C, etc.

4.1.1. Subject Recruitment and Diversity

The corpus was recorded by one hearing and two deaf signers, all proficient in Libras, which were instructed to dress as usual for the capture session. They also served as annotators to our data. To ensure that our data captures a wide range of physical constraints, recordings were conducted across three different vehicle models with varying cabin geometries: a Jeep Compass, a Nissan March, and a Renault Megane. All sessions were performed under clear daylight conditions to minimize extreme external lighting noise while preserving the natural dynamic shadows of the vehicle interior as seen in Figure 2.

(a) (b) (c)

Figure 2: Illustration of the three signers and challenges within vehicle cabin: (a) FLIR camera, (b) ToF camera, and (c) recording tablet point of views.

5. Data Acquisition and Technical Setup

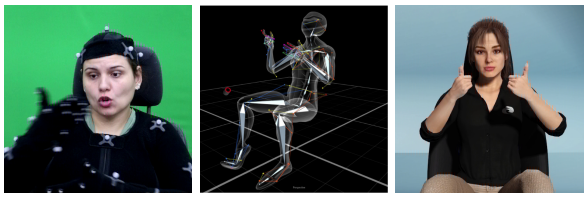
Our data collection followed a two-stage approach. We first established a laboratory baseline using high-precision motion capture to define the “gold standard” for each sign. This was followed by an acquisition phase inside a vehicle cabin, using a multimodal sensor setup to capture the natural constraints for in-car communication.

5.1. Laboratory MoCap Baseline

Our ICSL MoCap corpus was built to simulate in-vehicle body positioning and its effects on Libras sign production. To do so, we employed several sensors, including RGB cameras and a motion capture suit (MoCap), to record joint movements with detail and precision during sign language communication. The setup made use of several professionals, including:

- Two Libras signers who were fluent in Libras and performed the signs.
- One system operator to up-keep the quality of the MoCap data being recorded.
- Two hardware operators to ensure frame synchronization and camera operation, and lastly.
- One content facilitator to assist with translation and language production evaluation.

Regarding our frame composition for this task, the laboratory has a green-screen background and a mock-up car seat. The Libras’ signers are recorded while seated, wearing a special MoCap suit that covers most of their bodies. The suit is made from black cloth and allows the attachment of reflective beads (or spherical markers) with velcro, which are detectable by the infrared cameras as seen in Figure 3a. In total, there are 73 beads distributed across the main signer’s joints.



(a) (b) (c)

Figure 3: MoCap setup illustrating the (a) capturing suit, (b) the markerset and skeleton from the Vicon tracking software, and (c) the rendered avatar (Clara).

We also distributed fourteen Vicon (Vicon, 2026) cameras radially around the signer. These cameras are equipped with near-IR sensor that detect

the spherical markers attached to the MoCap suit. Later on the markers will be digitalized to move the character skeleton in the capture software as illustrated in Figure 3b, which in turn will animate the avatar as shown in Figure 3c. Moreover, two Canon EOS T6i Rebel (Canon, 2026) are also employed for RGB capture, alongside the FLIR 2D (Teledyne FLIR, 2026), and the Vicon Vue cameras.

5.2. In-Car Data Acquisition

The in-car component of the ICSL corpus was designed to capture the linguistic and environmental challenges of a transport system setting. We employed a multi-sensor approach to provide both standard video and high-fidelity 3D spatial data.

5.2.1. Hardware and Sensor Setup

We utilized multi-modal sensors as seen in Figure 4 and the setup as seen in Figure 5b was deployed across three distinct test vehicles: a Jeep Compass(illustrated in Figure 5a), a Nissan March, and a Renault Megane, providing variability in cabin dimensions, window size, lightning, color, texture, seat shape, and interior geometry.

(a) (b) (c)

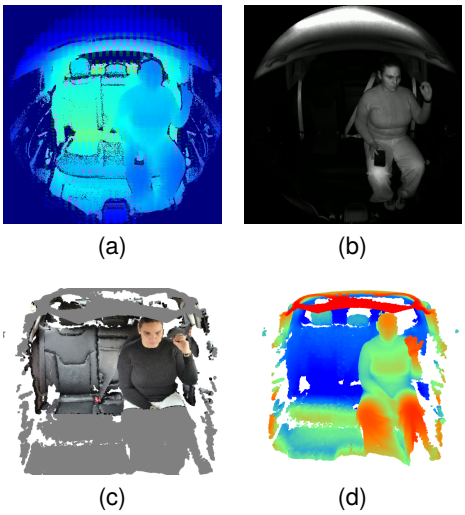
Figure 4: The multi-modal sensor suite including (a) the FLIR 2D camera, (b) Orbbec ToF sensor, and (c) the Lenovo Tablet.

- **2D Imaging:** A FLIR Blackfly S (Teledyne FLIR, 2026) was utilized to capture the primary 2D visual stream at a resolution of 1440×1080, operating at 60 fps as seen in Figure 4a. To accommodate the confined vehicle interior, we utilized fisheye lenses to maximize the field of view as seen in Figure 2a.
- **3D Time-of-Flight (ToF):** An Orbbec Femto Bolt (Orbbec, 2026) was utilized to capture spatial information as seen in Figure 4b. The sensor was configured to capture synchronized RGB, Depth, Infrared (IR), and Point Cloud streams at a resolution of 1024×1024 pixels, operating at 30 fps, allowing for precise tracking of hand movements in 3D space as seen in Figures 2b and 6a-d.
- **Recording Tablet:** A Lenovo Tab P12 (Lenovo, 2026) was utilized as an additional perspective, as seen in Figure 4c.

(a)

(b)

Figure 5: (a) test vehicle, and (b) full technical installation inside the Jeep Compass cabin, illustrating sensor placement and the passenger signing space.



(a) (b)

(c) (d)

Figure 6: Multimodal perspectives of Libras communication within a vehicle cabin, illustrating the two sensors: (a) ToF depth image shown in color, where blue indicates closer objects and green indicates farther objects. Dark blue areas mean no depth was measured, (b) ToF infrared (IR) image shown in grayscale, where white represents strong IR reflection and black represents weak or no reflection, (c) ToF RGB point cloud, where each 3D point contains spatial coordinates (X, Y, Z) and its corresponding color (R, G, B), and (d) ToF Depth point cloud, where each 3D point contains spatial coordinates (X, Y, Z) derived only from depth measurements, without color information.

Positioned behind the front-seat headrest at a resolution of up to 3840×2160 pixels, it recorded the signer from an elevated, centered viewpoint, providing a third data stream to supplement the mounted sensors as seen in Figure 2c.

- **Prompting Interface (Mobile Phones):** In addition to the main sensors, mobile phones were positioned within the signer's line of sight specifically to display the selected use cases. These served as a remote prompting system to guide the signer through the recording sequence.

5.3. Acquisition Protocol and Workflow

The recording session followed a strictly controlled protocol managed via a custom web-based control panel, as seen in Figure 7. To ensure the signer remained focused and in the correct posture, we utilized a remote signaling system:

- **Instruction Delivery:** Use cases were published on the mobile phone screens from the central Control Panel.
- **Start/Stop Signals:** The operator sent a "Start" signal (green) to the phone to indicate which sign to perform. Once the gesture was completed, a "Stop" signal (red) was sent, and the system automatically switched to the "Next Case" for the signer.

Figure 7: Web-based remote control interface for Start and Stop signaling.

6. Corpus Analysis and Limitations

Our dataset represents a first step toward a comprehensive in-vehicle Libras resource. To ensure the scientific validity of our comparative analysis, we chose to establish a solid baseline under stable, controlled conditions before introducing more dynamic environmental variables. This approach allows us to demonstrate the causal relationship between the physical layout of the car and signing production without the interference of visual noise. By focusing on these idealized conditions, we provide a "ground truth" reference that enables the evaluation of technical benchmarks in a confined space. This is critical because sign language representations must reflect 3D awareness to handle the orientations and spatial trajectories inherent in real-world cabins.

The current iteration of the dataset is built upon a theoretical vocabulary of 1,344 Portuguese context-specific phrases for ride-sharing applications. To date, a subset of 127 phrases has been successfully recorded in video, with plans to complete the remaining recordings in the coming months. The data collection and annotation methodology relies on a cohort of three participants (two deaf, one hearing) who perform both the signing and the ongoing data annotation.

A significant current limitation is the lack of dynamic lighting conditions, such as sun glare, tunnel transitions, or nighttime driving noise. While these factors are vital for "in-the-wild" robustness, in this initial phase, we theorize that models pre-trained on normal illumination data serve as a superior foundation for subsequent fine-tuning in low-light scenarios. As an early-stage resource within Project UNITY, the dataset currently features a limited number of signers, two deaf signer and one hearing. We acknowledge this as a potential source of signer bias that could limit the generalization of initial AI models. While we capture stationary occlusions (seatbelts, headrests), the dataset does not yet account for dynamic occlusions caused by vehicle movement or secondary passenger interactions.

By acknowledging these constraints, we frame this work not as a final solution but as the essential foundation required to guide our future, more complex data-collection efforts in the automotive sign language domain.

7. Conclusion

This paper introduces the In-Car Sign Language (ICSL) dataset, a novel multimodal resource designed to investigate the technical challenges of sign language recognition across shared mobility services. By focusing specifically on the constrained signing environment of vehicle interiors, we have established the first specialized dataset that systematically documents how Brazilian Sign Language (Libras) is produced within the physical and visual limitations of car cabins. The corpus comprises over 1.5 million synchronized frames across multiple modalities, providing researchers with a comprehensive foundation for investigating sign language recognition in real mobile settings.

Our contribution is split into two parts. First, a comparative benchmark between laboratory and in-car signing that addresses potential research direc-

tions in Section 3 focused on physical, multimodal, and linguistic adaptations. Second, we provide 127 communication scenarios recorded across three vehicle models, reflecting real-world accessibility needs within diverse cabin geometries.

The limitations acknowledged in Section 6 including controlled lighting conditions, and signer diversity, this work establishes a necessary foundation for robust "in-the-wild" SLR systems. As a component of Project UNITY, the ICSL corpus provides an empirical baseline for assistive technologies aiming to reduce communication barriers between hearing drivers and DHH passengers. Available upon request, this resource supports the development of inclusive transportation systems where language barriers no longer hinder DHH access and dignity.

8. Future Work

Moving forward, our next step is to introduce dynamic environmental parameters. We plan to record subsequent datasets under varied lighting conditions. Additionally, future iterations will move from a stationary vehicle to active driving environments to capture dynamic occlusions and noise caused by vehicle vibration, which are known bottlenecks for real-time landmark tracking. Although our dataset focuses on sentence-based communication scenarios, we aim to expand the ICSL corpus to include continuous Libras conversations between passengers, more complex phrases, and a broader range of subjects.

9. Acknowledgement

This project is partially financed by the São Paulo Research Foundation (FAPESP), grants #2024/23068-4 and #2024/00914-7, the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES), grant #88887.091672/2014-01, National Council for Scientific and Technological Development (CNPq), grant #458691/2013-5, and FINEP, grant #2778/20. The authors would also like to thank the Verkehrsverbund Großraum Ingolstadt (VGI) for their support through the VGI newMIND project.

10. Ethical Considerations

The Libras signers in this project are not just participants, they are paid team members and co-authors of this study. Deaf and hard-of-hearing researchers were involved in every stage of the project, from designing the data collection process to writing the paper. This helps us make sure that the linguistic details of Libras are treated with care and that the technology we develop truly reflects the needs of the community. All team members gave their informed consent for the use of their images and biometric data in the corpus. Since the data was created by the research team specifically for this resource, the project follows the institutional ethical guidelines for collaborative research. To request access to the corpus or for further technical inquiries regarding the acquisition software, annotation tools, or the Project UNITY framework, please contact José Mario De Martino at martino@unicamp.br.

11. Bibliographical References

Carlos Eduardo GR Alves, Francisco De A Boldt, and Thiago M Paixão. 2024. Enhancing Brazilian Sign Language recognition through skeleton image representation. In *37th SIBGRAPI Conference on Graphics, Patterns and Images*, pages 1–6. IEEE.

Sarah Alyami, Hamzah Luqman, and Mohammad Hammoudeh. 2024. Reviewing 25 years of continuous sign language recognition research: Advances, challenges, and prospects. *Information Processing & Management*, 61(5):103774.

Domenico Amalfitano, Vincenzo D'Angelo, Antonio Rinaldi, Cristiano Russo, and Cristian Tommasino. 2023. Enhancing gesture recognition for sign language interpretation in challenging environment conditions: A deep learning approach. In *15th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K)*, pages 395–402.

Jari Ivar Andersen, Gomer Otterspeer, Robert Belleman, and Floris Roelofsen. 2025. Designing a marker based motion capture setup for sign language research. In *Adjunct Proceedings of the 25th ACM International Conference on Intelligent Virtual Agents*, pages 1–9.

Necati Cihan Camgoz, Oscar Koller, Simon Hadfield, and Richard Bowden. 2020. Sign language Transformers: Joint end-to-end sign language recognition and translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Canon. 2026. EOS Rebel T6i product manual. https://www.cla.canon.com/en/p/eos-rebel-t6i. Digital camera technical documentation. Accessed: 2026-02-13.

Cleison Correia de Amorim, David Macêdo, and Cleber Zanchettin. 2019. Spatial-temporal graph convolutional networks for sign language recognition. In *International Conference on Artificial Neural Networks*, pages 646–657. Springer.

José Mario De Martino, Paula D Paro Costa, A Benetti, Luciana Aguera Rosa, Kate Mamhy Oliveira Kumada, and IR Silva. 2016. Building a Brazilian Portuguese–Brazilian Sign

Language parallel corpus using motion capture data. In *Proceedings of the 12th International Conference on the Computational Processing of the Portuguese Language, Tomar*, pages 56–63.

José Mario De Martino, Ivani Rodrigues Silva, Janice Gonçalves Temoteo Marques, Antonielle Cantarelli Martins, Enzo Telles Poeta, Dener Stassun Christinele, and João Pedro Araújo Ferreira Campos. 2025. Neural machine translation from text to sign language. *Universal Access in the Information Society*, 24(1):37–50.

Daniele LV dos Santos, Thiago B Pereira, Carlos Eduardo GR Alves, Richard JMG Tello, Francisco de A Boldt, and Thiago M Paixão. 2025. Proper body landmark subset enables more accurate and 5X faster recognition of isolated signs in LIBRAS. *arXiv preprint arXiv:2510.24887*.

Annika Herrmann. 2014. *Modal and Focus Particles in Sign Languages: A Cross-Linguistic Study*, volume 2. De Gruyter Mouton, Berlin, Boston.

Pavel Jedlička, Zdeněk Krňoul, Jakub Kanis, and Miloš Železnỳ. 2020. Sign language motion capture dataset for data-driven synthesis. In *9th Workshop on the Representation and Processing of Sign Languages*, pages 101–106.

Michael Kipp, Alexis Heloir, and Quan Nguyen. 2011. Sign language avatars: Animation and comprehensibility. In *International Workshop on Intelligent Virtual Agents*, pages 113–126. Springer.

Marie Lee, Ziming Li, Wendy Dannels, Tae Oh, and Roshan L Peiris. 2025. Exploring one handed signing during driving for interacting with in-vehicle systems for Deaf and hard of hearing drivers. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, New York, NY, USA. Association for Computing Machinery.

Lenovo. 2026. Tab P12: 12.7-inch Storm Grey tablet (8GB/128GB) product manual. https://www.mediamarkt.de/de/product/_lenovo-tab-p12pencil-8gb-128gb-127-zoll-tablet-128-gb-127-zoll-storm-grey-133566330.html. Hardware product specifications. Accessed: 2026-02-12.

Chenghong Lu, Misaki Kozakai, and Lei Jing. 2023. Sign language recognition with multimodal sensors and deep learning methods. *Electronics*, 12(23):4827.

Xiaohan Ma, Rize Jin, and Tae-Sun Chung. 2024. Multi-channel spatio-temporal transformer for sign language production. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 11699–11712.

Orbbec. 2026. Orbbec: Femto Bolt Time-of-Flight depth camera specifications. https://www.orbbec.com/products/tof-camera/femto-bolt/. Product technical datasheet. Accessed: 2026-04-07.

Presidência da República. 2002. Lei número 10.436. Diário Oficial da União, Imprensa Nacional, n. 79, Seção 1.

Reza Rastgoo. 2021. Sign language recognition: A deep survey. *Expert Systems with Applications*, 164:113794.

Dimitar Shterionov, Lorraine Leeson, and Andy Way. 2024. *The pipeline of sign language machine translation*, pages 1–25. Springer.

Hanan A. Taher and Subhi R. M. Zeebaree. 2025. A critical study of recent deep learning-based continuous sign language recognition. *The Review of Socionetwork Strategies*, 19(1):131–161.

Teledyne FLIR. 2026. Firefly S technical specifications. https://www.teledynevisionsolutions.com/products/firefly-s/?segment=iis&vertical=machine+vision. Machine vision camera documentation. Accessed: 2025-02-12.

Vicon. 2026. Vicon Valkyrie advanced motion capture (mocap) camera specifications. https://www.vicon.com/hardware/cameras/valkyrie/. Hardware technical datasheet. Accessed: 2026-02-13.

Ackley Dias Will, José Mario De Martino, and Juliano Renato S Bezerra. 2018. An optimized marker layout for 3D facial motion capture. In *Proceedings of the Smart Tools and Apps for Graphics (STAG)*, pages 107–113.

Wanyu Zhang. 2022. Sign language recognition based on depth image processing. *Highlights in Science, Engineering and Technology*, 23:25–33.

12. Language Resource References

Luis R. Cerna, E. E. Cardenas, D. G. Miranda, David Menotti, and Guillermo Camara-Chavez. 2021. *A Multimodal Libras-UFOP Brazilian Sign Language Dataset of Minimal Pairs using a Microsoft Kinect Sensor*. Expert Systems with Applications.

Ronice Müller de Quadros and Alexandre Melo de Sousa. 2021. *Corpus da Língua Brasileira de Sinais: Inventário de Libras do Acre*. Revista de Estudos da Linguagem.

M. A. Lima, D. Cruz, D. R. Silva, D. D. Albuquerque, D. F. Lacerda, R. Costa, G. L. de Souza Filho, and T. M. de Araújo. 2025. *VLibrasBD: A Brazilian Portuguese-Brazilian Sign Language (Libras) Bilingual Text Dataset*. Data in Brief.

Vicente Coelho Lobo-Neto and Helio Pedrini. 2024. *LSWH100: A Handshape Dataset for Brazilian Sign Language (Libras) using SignWriting*. Data in Brief.

Oline Ranum, Gomer Otterspeer, Jari I. Andersen, Robert G. Belleman, and Floris Roelofsen. 2024. *3D-LEX v1.0: 3D Lexicons for American Sign Language and Sign Language of the Netherlands*. arXiv preprint arXiv:2409.01901.

Tamires Martins Rezende, Sílvia Grasiella Moreira Almeida, and Frederico Gadelha Guimarães. 2021. *Development and Validation of a Brazilian Sign Language Database for Human Gesture Recognition*. Neural Computing and Applications.

Thad Starner, Sean Forbes, Matthew So, David Martin, Rohit Sridhar, Gururaj Deshpande, Sam Sepah, Sahir Shahryar, Khushi Bhardwaj, Tyler Kwok, Daksh Sehgal, Saad Hassan, Bill Neubauer, Sofia Vempala, Alec Tan, Jocelyn Heath, Unnathi Kumar, Priyanka Mosur, Tavenner Hall, Rajandeep Singh, Christopher Cui, Glenn Cameron, Sohier Dane, and Garrett Tanzer. 2023. *PopSign ASL v1.0: An Isolated American Sign Language Dataset Collected via Smartphones*. Curran Associates, Inc.