Co-funded by the
Erasmus+ Programme
of the European Union

ISENSE WP4
4.2 Motion data open database

Summary

Co-funded by the
Erasmus+ Programme
of the European Union

Introduction – ISENSE project

According to the World Health Organization (WHO), 466 million people worldwide suffer from hearing disabilities, with an estimated global cost of 178 billion euros per year due to social isolation and challenges in accessing quality education. This situation is particularly critical for university students with hearing impairments, who require specific support to navigate the academic path. This is the driving force behind the ISENSE project (Innovative Supporting Services for University Students with Deafness), which aims to create an inclusive educational ecosystem that allows deaf students to achieve full autonomy during their academic journey and integrate into the workforce without facing isolation compared to their peers. The main goals of the project are to reduce educational barriers, train university staff, and develop innovative technologies.

To achieve these goals, ISENSE employs an approach based on immersive technologies and artificial intelligence (AI).

For more details, visit the website www.isenseproject.eu

Document aim

This is the document containing all the details related to the database realized within SHARE SLANG (WP4), containing motion data related to the Italian sign language. Through this document, research groups can understand how the data has been acquired and how to use the database for their own application.

Protocol 1 - Sensor system

The sensor system used for the acquisition of motion data is Rokoko. Rokoko is recognized as a leading company in the field of human movement digitization and motion capture technologies. Founded in 2014 in Copenhagen, Denmark, the company quickly established itself as a leader in 3D animation and motion capture technology. Rokoko's primary goal is to make motion capture

accessible to a broad and diverse audience, spanning industries such as entertainment, education, and research.

The Rokoko products, which will be described below, were used in this study to assess whether they could be employed in the creation of a database for the Italian Sign Language (LIS) and other sign languages.

Rokoko – SmartSuit Pro II

The Rokoko Smartsuit Pro II is an advanced motion capture suit designed to record and analyze human movements in real time with high precision. Made from lightweight, flexible material, it covers the entire body, offering comfort and freedom of movement. Equipped with inertial sensors providing 9 degrees of freedom, the suit measures orientation, speed, and acceleration. The electronic sensors are removable, allowing for the suit to be washed or adjusted to fit different sizes. With a 200 fps refresh rate and wireless connectivity, the suit ensures smooth, delay-free streaming without the need for cameras or dedicated tracking spaces (Figure 1).

Figure 1 SmartSuit Pro II by www.rokoko.com

Rokoko – SmartGloves

The Rokoko Smartgloves are gloves designed to accurately monitor and record hand and finger movements in real time, ideal for motion capture and interaction in virtual environments. They can be used independently or in combination with the Smartsuit, with Wi-Fi connectivity and a range of

up to 100 meters. Each glove is equipped with 7 inertial and EMF sensors, with a 100 fps refresh rate, ensuring lag-free recordings (Figure 2).

Figure 2 Smartgloves by www.rokoko.com

Protocol 1 - Data

Experimental protocol

Data in the database have been acquired following a specific experimental protocol, here depicted. The study was conducted by acquiring only the letters of the Italian Sign Language (LIS) alphabet (Figure 3). The manual alphabet is used to represent words, names, or technical terms for which there is no codified gesture in sign language, allowing for more precise communication in situations where existing signs are insufficient.

The hand positions associated with each letter of the alphabet are characterized by specific movements and the use of particular finger and wrist joints. For example, the representation of the letter "A" requires the hand to be clenched into a fist, with the fingers bent and the thumb resting on the side of the fingers. In contrast, the letter "B" requires the hand to be open, with the fingers extended and together, while the wrist remains in a neutral position.

The letters also differ in terms of the dynamism of their execution. There are static letters, which include: A, B, C, D, E, F, H, I, L, M, N, O, P, Q, R, S, T, U, V, W, X, and Y. In contrast, dynamic

letters are characterized by a movement of the wrist or forearm during execution, and these include G, J, and Z.

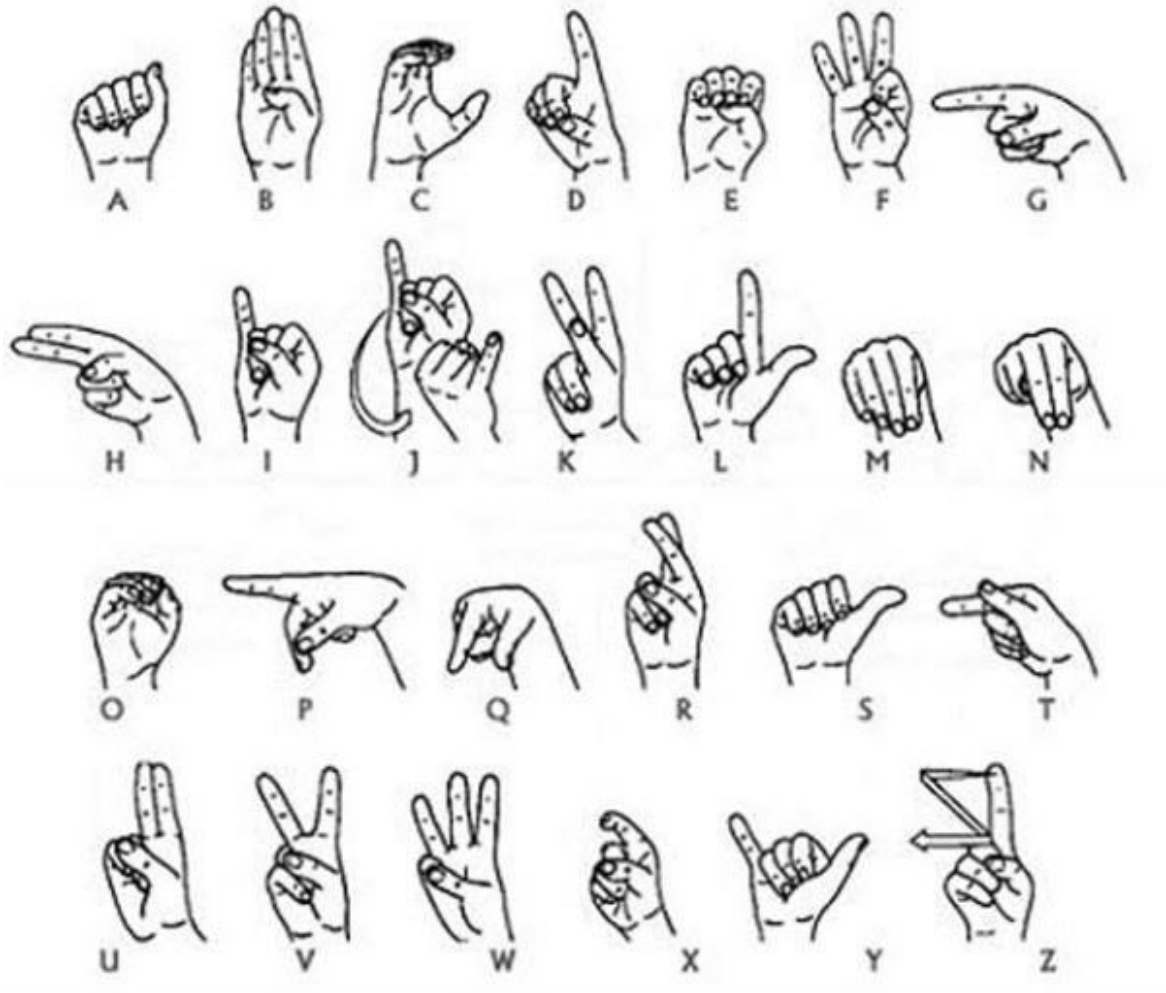

Figure 3 Letters of the LIS alphabet

Data gathered from ten participants were included in the database. Among the selected participants, eight had no prior knowledge of Italian Sign Language (LIS). For these participants, a familiarization phase was conducted before data acquisition to provide awareness of the gestures to be performed, reducing errors in movement reproduction. The remaining two participants had significant experience with LIS. To minimize variations due to external factors and ensure uniform experimental conditions, all participants were selected based on specific criteria, such as wearing the same suit size (M) and

being right-hand dominant. These homogeneity criteria were aimed at minimizing differences related to suit fit and hand dominance.

Figure 3 Three of the subjects acquired

Acquisition

During the movement recording, participants were asked to wear the Smartsuit and Smartgloves. After setting up the devices, participants were instructed to stand in the center of the room to perform a calibration, which was a preliminary step before acquiring the movement data of the letters. Once calibration was completed, the protocol to be followed was thoroughly explained, including instructions for performing the gestures related to each letter. The movement recording was then initiated, with the participant starting from the predefined position according to the experimental protocol (with the dominant hand raised and open), executing the gesture corresponding to the letter of interest, and then returning to the initial position. This process was repeated eight times for each letter, with the recording being stopped when the participant returned to the starting position.

At the end of the recording for a letter, the participant returned to the calibration position, and the procedure described was repeated for each letter of the alphabet. After completing the acquisition of all twenty-six letters, the participant removed the Smartgloves and then the Smartsuit, before putting them on again. The procedure was repeated for a total of three sessions for each participant.

Data is acquired with a sampling frequency of 30 Hz.

Export

The data acquired using Rokoko software was exported in BVH format and subsequently opened in the 3D modeling environment Blender. In this software, the files automatically adopt Blender's

absolute reference system, centered on the "Hips" segment (Figure 4). This facilitated the management of movement data, as Blender provides a clear and well-defined reference system, allowing for an accurate understanding of the direction of movement of each joint along the three-dimensional axes. Data has been converted into a CSV file.

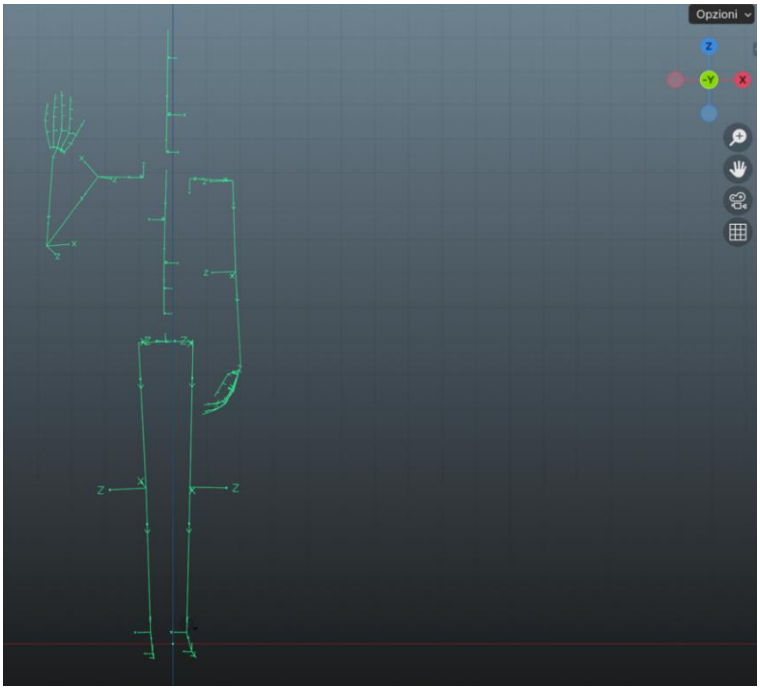

Figure 4 Reference system Blender

Structure

The CSV files are stored in the folder named “LIS alphabet”. This database is open access and the data can be downloaded by all people.

Inside this folder, the CSV files are named according to the following convention:

$$P_i \_ S_j \_ L_k \, . csv$$

where:

- **i** represents the participants, ranging from 1 to 10; thus, P_i is the *i-th* participant;
- **j** represents the session number, ranging from 1 to 3; thus, S_j is the *j-th* session;
- **k** represents the letter; thus, L_k is the *k-th* letter of the LIS alphabet.

The CSV file is organized to record data related to the movements of the bones in the 3D model, with particular focus on the rotation and translation of each bone. Each row in the file corresponds to an acquisition frame and contains detailed information about the position and orientation of the bones according to the relative frame. In the Figure 5 below, you can see a portion of the output CSV file, where commas between values denote the separation of columns.

```
Frame,Bone,Rotation Matrix 11,Rotation Matrix 12,Rotation Matrix 13,Rotation Matrix 14,Rotation Matrix 21,Rotation Matrix 22
1,Hips,1.0,0.0,0.0,0.0,0.0,0.42193588614463806,-0.9066256880760193,6.73594513500575e-06,0.0,0.9066256880760193,0.42
1,Spine1,1.0,-1.5099553252184705e-07,-1.509951914613339e-07,0.0,-1.5099600148005266e-07,-1.7453299960834556e-06,
1,Spine2,1.0,-1.5099553252184705e-07,6.394884621840902e-13,-1.140016706813185e-06,1.4210841162674848e-14,-1.780
1,Spine3,1.0,-1.5099553252184705e-07,7.927983356609269e-13,-2.2800331862526946e-06,1.705302701349512e-13,-1.760
1,Spine4,-1.0,-1.5045510792788264e-07,-1.6375928169054532e-07,-4.2050296542583965e-06,-1.5045577583805425e-07,-0
1,Neck,1.0,-1.5099546146757348e-07,8.100187187665142e-13,-7.247634130180813e-06,1.5631940186722204e-13,-1.75088
1,Head,1.0,-1.5099546146757348e-07,7.952380615995625e-13,-8.923819223127794e-06,1.4210854715202004e-13,-1.74711
1,HeadTip,1.0,-1.5099546146757348e-07,7.973741025096848e-13,-1.1468637239886448e-05,1.4210853359949288e-13,-1.7
1,RightShoulder,-1.8763995512927067e-06,-1.0,4.694578059627474e-13,-7.000006198883057,-1.7439758721593535e-06,3.1
1,RightArm,1.2115136449608599e-08,-0.9999999403953552,1.8636368395164027e-06,-20.607803344726562,-1.0000011920
1,RightForeArm,-1.7332134802927612e-06,-0.9999999403953552,1.8602858062877203e-06,-47.71000671386719,-1.000001
1,RightHand,0.22192512452602386,-0.9742657542228699,-0.03943741321563721,-76.010009765625,-0.9138233065605164,
1,RightFinger1Metacarpal,0.3535546362400055,-0.8660272359848022,-0.3535475432872772,-78.08802032470703,-0.612375
1,RightFinger1Proximal,8.810311555862427e-07,-0.9999999403953552,6.288290023803711e-06,-81.54269409179688,-0.707
1,RightFinger1Distal,8.810311555862427e-07,-0.9999998807907104,6.278473847487476e-06,-84.57049560546875,-0.70710
1,RightFinger1Tip,8.810311555862427e-07,-0.9999998807907104,1.0424942047393415e-06,-88.02949523925781,-0.7071096
1,RightFinger5Metacarpal,-0.342021644115448,-0.939691960811615,3.713314526976319e-06,-78.49790954589844,-0.93969
1,RightFinger5Proximal,-1.7583348608241067e-06,-0.9999998807907104,1.955120751517825e-06,-82.95796966552734,-1.0
1,RightFinger5Medial,-1.7583348608241067e-06,-0.9999998807907104,1.9411872926866636e-06,-86.45796966552734,-1.00
1,RightFinger5Distal,-1.7583348608241067e-06,-0.9999998211860657,1.9933952444262104e-06,-88.65797424316406,-1.000
1,RightFinger5Tip,-1.7583348608241067e-06,-0.9999997615814209,1.993390469579026e-06,-90.89887237548828,-1.000001
1,RightFinger4Metacarpal,-0.1736498326063156,-0.9848072528839111,3.646214963737293e-06,-78.75300598144531,-0.984
1,RightFinger4Proximal,-1.7285409512624028e-06,-0.9999997615814209,3.309418843855383e-06,-83.70904541015625,-1.0
```

Figure 5 CSV file structure

Below is a detailed description of what each row and column of the file contains:

1. First Row: File Header

The first row contains the column names that describe the data in the file. The columns are structured as follows:

- **Frame**: Indicates the frame number (e.g., 1, 2, 3, etc.), useful for identifying the time stamp.
- **Bone**: Name of the bone (e.g., "Hips," "Spine1," etc.), representing the part of the body for which data is recorded.
- **Rotation Matrix (11-44)**: These 16 values represent the full roto-translation matrix of the body segment, which includes both rotation and translation. This matrix, obtained using the rotation_matrix function, is a 4x4 matrix that consists of the rotational component (3x3) and the translational component (1x3 in the fourth column). Each value in this matrix is identified by a column using the following nomenclature:

  $$RotationMatrix\ mn$$

  where:

  - **m** represents the row number of the rotation matrix.
  - **n** represents the column number of the rotation matrix.

  Rotation Matrix 11,Rotation Matrix 12,Rotation Matrix 13,Rotation Matrix 14

- **Translation Vector X, Y, Z:** These three values represent the absolute translation vector, which describes the translation of the segment relative to the origin of the global coordinate system. Each value of this vector is identified by a column with the following naming convention:

  Translation Vector X,Translation Vector Y,Translation Vector Z

- **Relative Rotation Matrix (11-44):** These 16 values describe the relative roto-translation matrix with respect to the parent segment. This matrix, obtained using the relative_matrix function, is also 4x4 in size and describes the transformation of the segment relative to its parent segment. The relative matrix is calculated by multiplying the inverse matrix of the parent segment by the matrix of the current segment, ensuring an accurate representation of the relationship between the segments in the armature hierarchy. Like the rotation matrix, it follows the same naming convention as mentioned above, with the term "Relative" prefixed before "Rotation Matrix."

- **Relative Translation Vector X, Y, Z:** These three values represent the relative translation vector, obtained using the relative_translation_vector function, which describes the translation of the segment relative to its parent.

2. Second Row: Data for the First Frame for the Bone "Hips"

The second row contains the actual data for the first frame (Frame 1) related to the bone "Hips." The structure of the data follows the header and includes the following values:

- **Frame**: 1 (indicates the first frame).
- **Bone**: Hips (the reference bone).
- **Rotation Matrix (11-44)**: The global rotation matrix of the bone "Hips," expressed through 16 values.
- **Translation Vector X, Y, Z**: Indicate the position of the bone in global space.
- **Relative Rotation Matrix (11-44)**: The rotation matrix related to the previous bone.
- **Relative Translation Vector X, Y, Z**: Identify the relative position of the bone compared to the previous bone.

3. Third Row: Data for the First Frame for the Bone "Spine1"

The third row follows the same structure but pertains to the bone "Spine1." The values are organized in the same order but differ for the specific data related to the orientation and position of this bone:

- **Frame**: 1 (still the first frame).
- **Bone**: Spine1.
- **Rotation Matrix (11-44)**: The global rotation matrix for the bone "Spine1."
- **Translation Vector X, Y, Z**: Position of the bone "Spine1" in global space.
- **Relative Rotation Matrix (11-44)**: Rotation matrix relative to the previous bone.
- **Relative Translation Vector X, Y, Z**: Relative position compared to the previous bone.

4. Subsequent Rows

Each subsequent row follows the same structure for each bone and for all recorded frames. The format remains consistent: first the global orientation and position, followed by the relative values. All the bones present within the parent file are shown in the image below Figure 6, which also includes the

parent bone for global reference, as well as the hierarchy used to calculate the relative rotation matrix between the considered bone and the previous one.

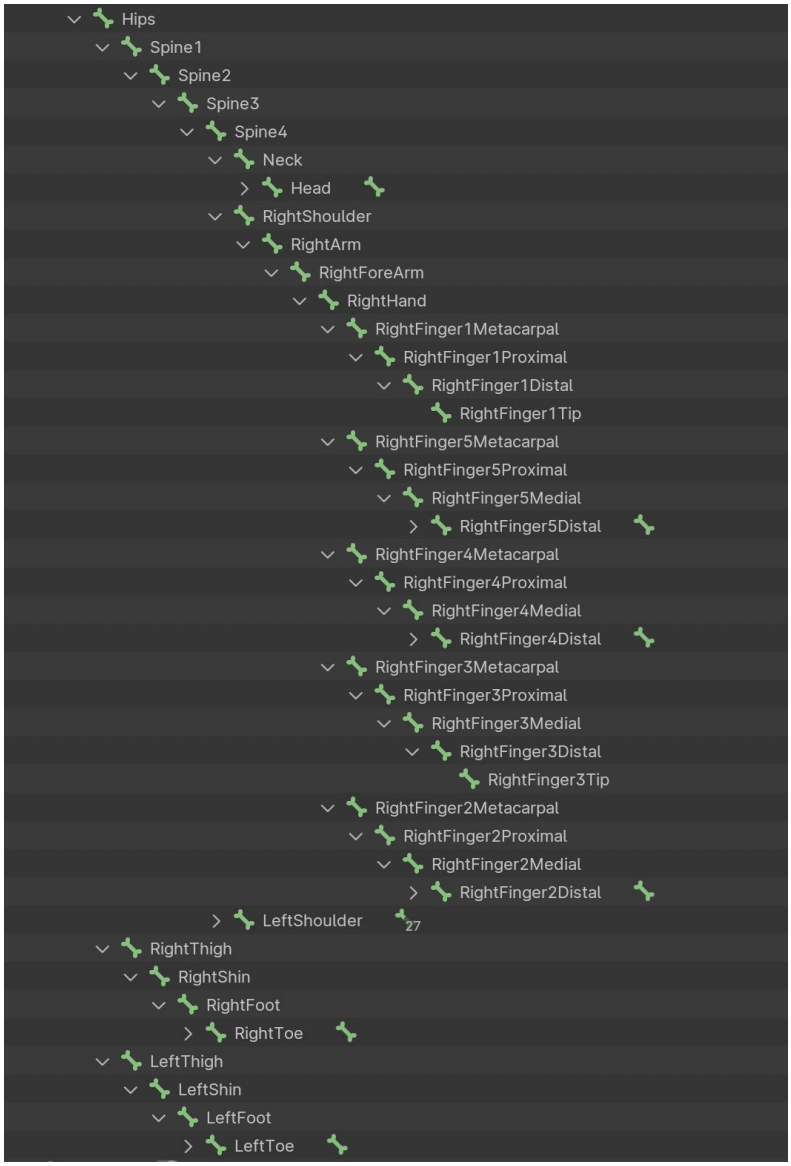

Figure 6 Segments included in the database with relative parent

Data analysis

This paragraph only shows a possible flowchart for the data analysis; however, the database is populated with raw data (.csv file), like described above.

The analysis of the data was carried out using a script developed in MATLAB, specifically designed to process the information contained in a CSV file. This CSV file contains essential details regarding the rotations and positions of the body segments of the arm and fingers. The data imported from the CSV file was processed using the readtable function, which allows for reading and importing data into a structured table. In the resulting table (Figure 7), the first column represents the frames

of the acquisition, whereas the second column identifies the body segments ('bones') considered by Blender during the conversion from BVH to CSV. The rows of the table contain the rotation matrices and the corresponding values for each reference 'bone'.

	1 Frame	2 Bone	3 RotationMatrix11	4 RotationMatrix12	5 RotationMatrix13
1	1	'Hips'	1	0	0
2	1	'Spine1'	1	-1.5100e-07	-1.5100e-07
3	1	'Spine2'	1	-1.5100e-07	6.5370e-13
4	1	'Spine3'	1	-1.5100e-07	9.0648e-13
5	1	'Spine4'	-1	-1.5046e-07	-1.6376e-07
6	1	'Neck'	1	-1.5100e-07	8.9528e-13
7	1	'Head'	1	-1.5100e-07	8.7938e-13
8	1	'HeadTip'	1	-1.5100e-07	8.8151e-13
9	1	'RightShoulder'	-1.8764e-06	-1	3.8392e-13
10	1	'RightArm'	1.2115e-08	-1	1.8636e-06
11	1	'RightForeArm'	-1.7332e-06	-1	1.8519e-06
12	1	'RightHand'	0.2219	-0.9743	-0.0394
13	1	'RightFinger1Metacarpal'	0.3536	-0.8660	-0.3535
14	1	'RightFinger1Proximal'	7.7300e-07	-1.0000	6.1095e-06
15	1	'RightFinger1Distal'	7.7300e-07	-1.0000	6.0997e-06
16	1	'RightFinger1Tip'	7.7300e-07	-1.0000	8.6368e-07
17	1	'RightFinger5Metacarpal'	-0.3420	-0.9397	3.7777e-06
18	1	'RightFinger5Proximal'	-1.7583e-06	-1	1.9481e-06
19	1	'RightFinger5Medial'	-1.7583e-06	-1	1.9342e-06
20	1	'RightFinger5Distal'	-1.7583e-06	-1	1.9864e-06

Figure 7 Table CSV data extraction

For each body segment, the script extracts the rotation matrices for each frame, both relative to the previous segment and to the "Hips" (parent bone). This extraction process is repeated iteratively for each acquired frame. At the end of the processing, all rotation matrices are stored in a data structure called a "map."

Considering that the study focuses exclusively on the letters of the Italian Sign Language (LIS) and that these letters require the exclusive use of the upper limbs, it was necessary to filter the data to include only the segments of interest. To this end, a dedicated folder for the joints of the upper limbs was created, excluding other body parts.

The calculation of joint rotations was performed by multiplying the transposed rotation matrix of the previous segment—referred to the global coordinate system "HIPS"—with the rotation matrix of the subsequent segment. This approach, used in biomechanical analysis, is essential for "nullifying" the effects of the accumulated rotations up to that point, isolating the actual rotation of the joint between the two segments and focusing on the relative movement between them. In other

words, transposing the rotation matrix of the previous segment allows for the measurement of the joint's rotation independently.

$$R_{sp}^{sd} = (R_{sd}^{0})^{T} \; X \; R_{sp}^{0}$$

- $(R_{sd}^{0})^{T}$ represents the transposed rotation matrix of the previous segment with respect to the absolute reference system "Hips" (0).
- R_{sp}^{0} is the rotation matrix of the subsequent segment with respect to the absolute reference system.
- R_{sp}^{sd} is the relative rotation of the joint located between the two segments.

For each joint, for which the rotation between the two segments has been calculated, a name based on the anatomy of the hand and arm has been assigned. Specifically, for the arm, considering the segments exported from the CSV (Figure 8), the following joints have been identified:

- **Joint Shoulder** (Shoulder)
- **Joint Arm** (Elbow)
- **Joint Hand** (Wrist)

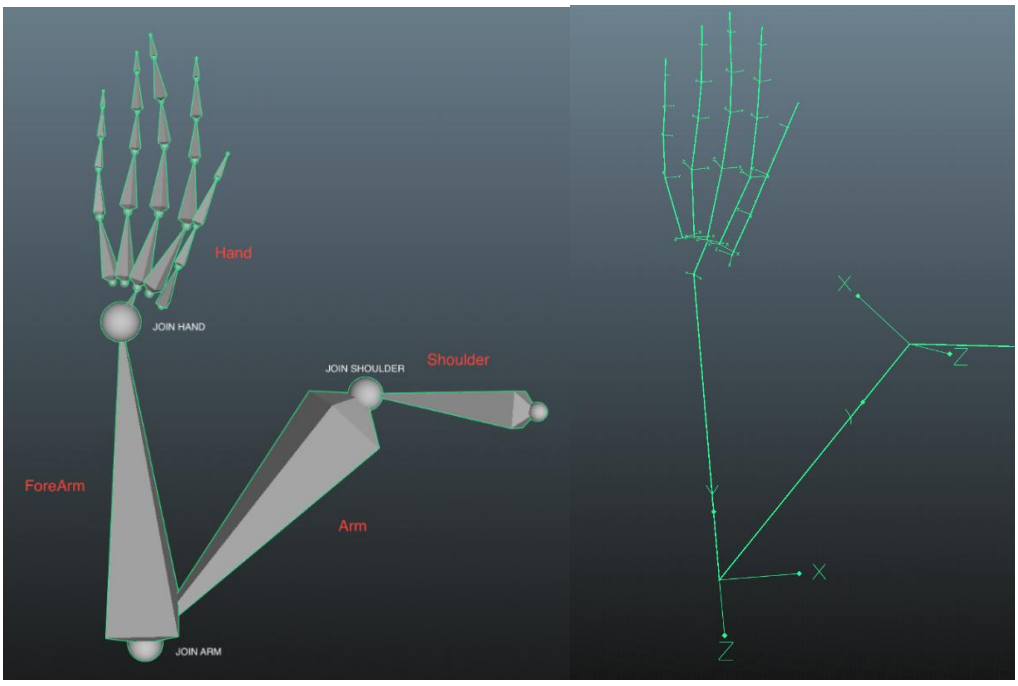

Figure 8 a) Arm segment in CSV file b) Reference system in Blender

For the hand, also considering the segments exported from the CSV (Figure 9), the following joints have been identified:

- **Joint Metacarpal** (Metacarpal)
- **Joint Proximal** (Proximal)
- **Joint Medial** (Medial)
- **Joint Distal** (Distal)

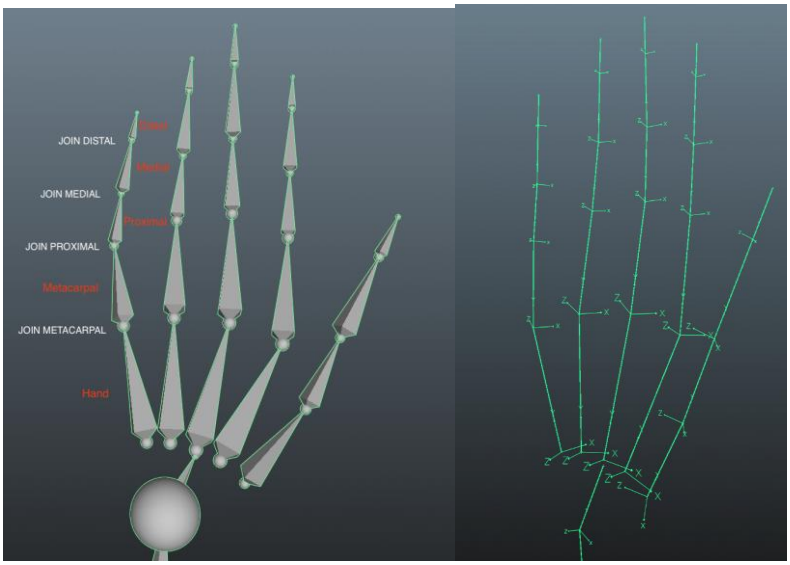

Figure 9 segment and joint hand's finger

Co-funded by the
Erasmus+ Programme
of the European Union

Protocol 2 – Aim

Create a multimodal database of Italian Sign Language (LIS) in a university context. The dataset will include 14 expert participants (deaf signers, LIS interpreters, or other LIS experts). Each participant will produce 20 target words and 5 sentences all related to the University context. Each word and sentence is then repeated five times.

Protocol 2 – Setup

An optoelectronic system plus two HD cameras and a face-capture system. For the optoelectronic system, it is required the use of full upper limbs model combined with the kinematics of trunk, hand and head.

Protocol 2 – List of words

1. **student** – studente
2. **professor** – professore
3. **university** – università
4. **classroom** – aula
5. **library** – biblioteca
6. **office** – ufficio
7. **Erasmus** – erasmus
8. **enrollment/registration** – immatricolazione
9. **ID card / badge** – tesserino
10. **schedule / timetable** – orario
11. **exam** – esame
12. **grade / mark** – voto
13. **scholarship** – borsa di studio
14. **tuition / fee** – tassa universitaria
15. **canteen / cafeteria** – mensa
16. **residence / dormitory** – residenza
17. **orientation** – orientamento
18. **information desk** – sportello informazioni
19. **password / credentials** – credenziali
20. **internet / Wi-Fi** – Wi-Fi

Protocol 2 - List of sentences

1. **I am looking for the orientation office.**
   Sto cercando l'ufficio orientamento.
2. **What is the procedure for enrollment?**
   Qual è la procedura di immatricolazione?
3. **Is there an entrance test?**
   C'è un test di ingresso?
4. **What courses of study does the university offer?**
   Quali corsi di studio offre l'università?
5. **Where is the classroom?**
   Dove è l'aula?

Useful link

Optoelectronic data:

https://drive.google.com/drive/folders/15TUGcIuxcHlXP2UGr8pedM7JFd8CDCzj?usp=drive_link

This link contains the direct access to the database related to Procol2, concerning the acquisition made by optoelectronic system.

Video-camera data:

https://drive.google.com/drive/folders/1lBbHu45T33KJJyFkCXLwZkRdD7m3cj2r?usp=drive_link

This link contains the direct access to the database related to Protocol 2, concerning the acquisition made by the two cameras.

Inertial system data:

https://drive.google.com/drive/folders/189qCAbhgIzlZi-TQG7kBMrnYFnl9kI1b?usp=drive_link

This link contains the direct access to the database related to Protocol 1, concerning the acquisition made by inertial measurement units embedded in Rokoko System.

Facial expression:

https://drive.google.com/drive/folders/16ptOt5vle3QEoool4jnjnf_7OqmM8uwV?usp=drive_link

This link contains the direct access to the database related to Protocol 2, concerning the acquisition made by ArKIT technology

For all information about the open access use of database please contact mibs@unitus.it

Conclusions

This document provides a detailed overview of the data structure within the database dedicated to the Italian Sign Language (LIS) alphabet. The organized and comprehensive nature of this dataset allows for effective cataloging and retrieval of key information relevant to each sign in the alphabet. By standardizing the representation of LIS data, this resource aims to support researchers and developers, particularly those working in fields like machine learning, computer vision, and natural language processing. Future applications could leverage this database to enhance models for sign language recognition, automated translation, or interactive educational tools, thus contributing to advancements in accessibility and inclusivity for the Deaf community.