GREEK SIGN LANGUAGE RECOGNITION FOR THE SL-REDU LEARNING PLATFORM

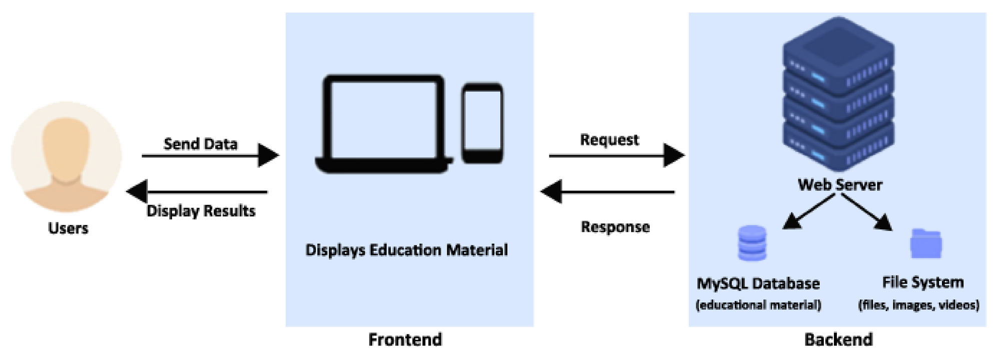
Katerina Papadimitriou [1] Gerasimos Potamianos [1] Galini Sapountzaki [2]
Theodor Goulas [3] Eleni Efthimiou [3] Stavroula-Evita Fotinea [3] Petros Maragos [4]

[1] Dept. of Electrical & Computer Eng., University of Thessaly, Volos, Greece
[2] Dept. of Special Education, University of Thessaly, Volos, Greece
[3] Institute for Language & Speech Processing, Athena RC, Athens, Greece
[4] School of Electrical & Computer Eng., National Technical University of Athens, Greece

Overview

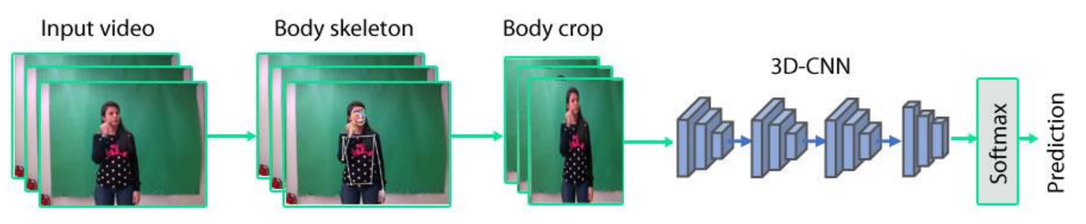
- **Goal:**
  - ✓ **SL education tool** development for SL production **self-assessment** and **objective evaluation**.
  - ✓ **SL recognition (SLR)** from videos in a **signer-independent (SI) mode** under **realistic recording conditions**.
- **Previous work [1, 2, 3]:**
  - ✓ A **suitable platform** for the **SL-ReDu project** is built involving "passive"-type GSL learning exercises.
- **SL recognizer:**
  - ✓ **Recognition module** based on state-of-the-art **deep-learning** techniques.
  - ✓ Focus on **isolated signs** and **continuously fingerspelled** letter sequences.
- **Contributions:**
  - ✓ **Early version** of a **GSL recognizer** integrated within the **SL-ReDu learning platform**.
  - ✓ **First** education tool in **GSL** with **recognition functionality**.
- **Evaluation:**
  - ✓ **High multi-signer (MS)** and **SI** recognition **accuracies**.
  - ✓ **Evaluation** by student and expert users of the **SL-ReDu platform** and its **recognition functionality** demonstrates very **satisfactory** objective and subjective **assessments**.

SL-ReDu Platform

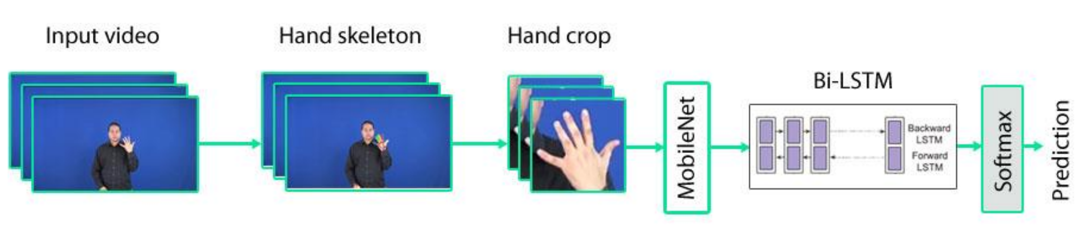


- **Enables self-monitoring and objective learner evaluation.**
- **System's design involves all aspects of GSL linguistics:**
  - ✓ Teaching **techniques** and **content**, including various **SL practice assignments**.
  - ✓ **Multiple-choice questions:** images, videos, and text.
  - ✓ **User response** or **user feedback** by means of **video recordings** of **GSL production**.
  - ✓ **Enables** the user to **actively sign** and be **assessed** for the capacity to appropriately generate signs.
- **SL-ReDu prototype system:**
  - ✓ **Web-based** application managing the **end user's interaction**.
  - ✓ **System** modalities entail the system **database**, the **front-end** and **back-end** user interfaces, and image/video **files**.
  - ✓ **SLR** is a separate system module **running** as standalone on the **learner's device**.

GSL Recognition Module



(Isolated sign recognition)

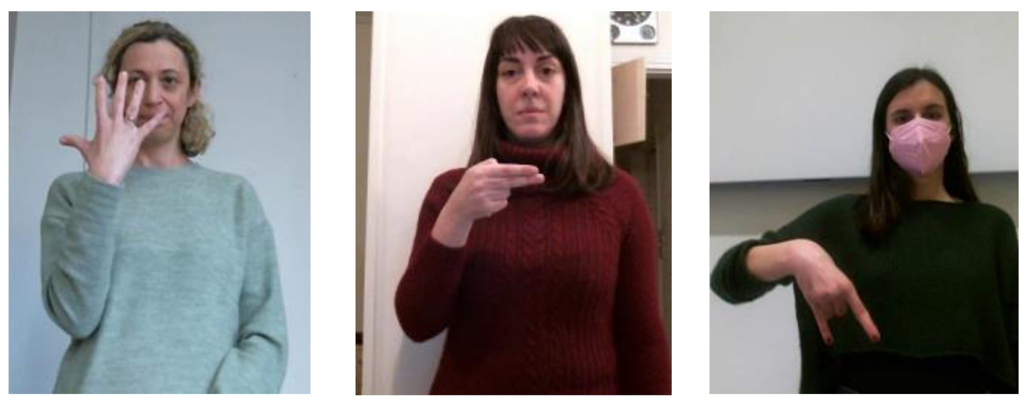


(Continuous fingerspelling recognition)

- **Pre-processing:**
  - ✓ **Detect** the **signer**, **extract** the Region-of-Interest (**RoI**), and **provide feedback** in case of incorrect **signer positioning**.
  - ✓ **MediaPipe** library for signer's **whole-body landmarks extraction** from **RGB video**.
  - ✓ **Lack** of detected **landmarks** of hands, face, and upper torso: **incorrect** user **positioning**.
  - ✓ **Correct** user **positioning: RoI extraction**.
  - ✓ **Isolated sign**: **upper body** is cropped producing the **RoI**.
  - ✓ **Fingerspelling**: **RoI** consists of the **signing hand**.
- **Isolated Sign Recognition:**
  - ✓ 18-layer **ResNet2+1D** model is used, separating **spatial** and **temporal convolutions** of 3D CNNs.
  - ✓ **Two** recognition **subtasks** employing **separate models**:
    - **Numeral signs** with a vocabulary size of 18.
    - **Non-numeral signs** with a vocabulary size of 36.
- **Continuous Fingerspelling Recognition:**
  - ✓ **CNN-BiLSTM** model is used.
  - ✓ **MobileNet** based visual **feature learner** of each video frame: 1024-dim features.
  - ✓ **Features** are fed to a **linear projection** layer for **size reduction:** 512-dim features.
  - ✓ Two-layer **BiLSTM encoder** followed by **CTC decoding**.

GSL Database

- **Signing data by multiple volunteer informants both native and non-native in GSL:**
  - ✓ Data **recorded indoors**, under realistic, **non-studio conditions**.
- **Numeral signs:**
  - ✓ **20** signers **x 18** signs **x 5 times: 1,800 videos**.
- **Non-numeral signs:**
  - ✓ **17** signers **x 36** signs **x 5 times**.
  - ✓ **ITI GSL corpus [4]: 7** signers **x 36** signs **x 5 times**.
  - ✓ **Total: 24 (17 + 7)** signers - 4,320 videos.
- **Fingerspelling data:**
  - ✓ **12 informants:** 24 Greek **alphabet letters** and **50 fingerspelled words** of 4-5 letters (unique to each signer).
  - ✓ **7 informants: 16** 3-7 letter **words** (common to all).
  - ✓ **3 signers** expressed **extra 71 words** of 4-5 letters.
  - ✓ **Total**: **1,071 videos**.



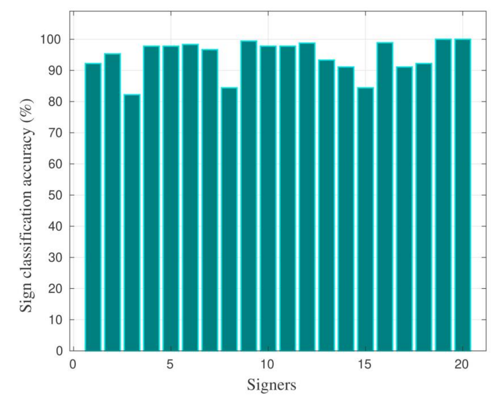
(Numerals) (Non-numerals) (Fingerspelling)

Experimental Framework

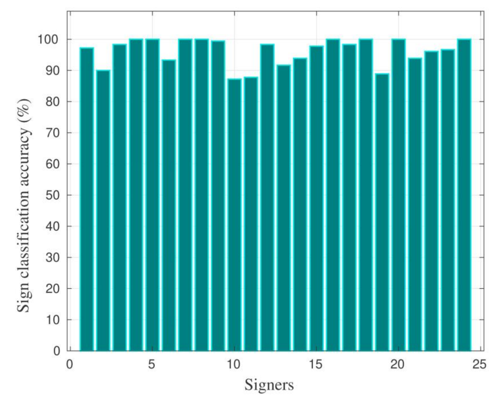
- **Multi-signer (MS) recognition:**
  - ✓ **Training 80%** of all **videos** (numerals: 1,440; non-numerals: 3,456; fingerspelling: 857).
  - ✓ **Validation 10%** of all **videos** (numerals: 180; non-numerals: 432; fingerspelling: 107).
  - ✓ **Testing 10%** of all **videos** (numerals: 180; non-numerals: 432; fingerspelling: 107).
- **Signer-independent (SI) recognition:**
  - ✓ **20-fold** cross-validation for **numerals**.
  - ✓ **24-folds** cross-validation for **non-numerals**.
  - ✓ **12-folds** cross-validation for **fingerspelling**.
  - ✓ **Each fold** contains **one test subject**, all **remaining subjects** are used in **training**.
- **SL-ReDu platform user evaluation:**
  - ✓ **Training**: **90%** of the available **videos** (numerals: 1,620; non-numerals: 3,888; fingerspelling: 964).
  - ✓ **Validation**: **10%** of the available **videos** (numerals: 180; non-numerals: 432; fingerspelling: 107).
- **GSL recognition performance:**
  - ✓ **Isolated** GSL and **continuous** fingerspelling tasks under both **MS** and **SI** training/testing cases.
  - ✓ GSL recognizer **objective evaluation** results.
  - ✓ **Results** in word accuracy (**WAcc**) %, and in the case of fingerspelling in letter accuracy (**LAcc**) %.
  - ✓ **Isolated GSL recognition task:**
    - **Performance degrades** in the **SI** case.
    - **WAcc satisfactory** in **both isolated** SLR **tasks**.
    - **Objective evaluation**: **results better** than **SI** case.
  - ✓ **Continuous fingerspelling recognition task:**
    - **Performance suffers** at the **WAcc** level, especially for longer letter sequences.
    - **Higher LAcc** results.
    - **Objective evaluation**: **results better** than **SI** scenario.

GSL recog. task	Metric	MS	SI	Eval.
iso. numerals	WAcc	97.78	94.48	98.61
iso. non-numerals	WAcc	99.44	96.20	97.22
cont. fingerspelling	WAcc	75.22	65.30	90.28
	LAcc	86.12	77.66	91.03

- ✓ **SI isolated GSL recognition** accuracy (%) per signer for **numerals** and **non-numeral** signs.
  - **Performance varies** among signers, **remaining** nevertheless well above **80% Wacc**.



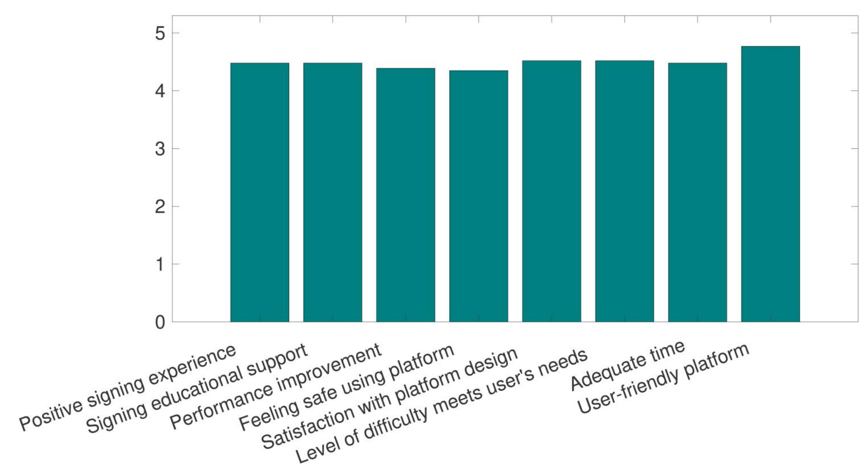
Numerals



Non-numerals

SL-ReDu Platform User Evaluation

- **Volunteer Users:**
  - ✓ **Department of Special Education at University of Thessaly students:**
    - **Group 1**: 10 students (GSL for less than 5 months).
    - **Group 2**: 11 students (GSL for more than 5 months).
  - ✓ **Two GSL experts/teachers volunteers**.
  - ✓ **Ages: 19-22** years old / **Females**> **males**.
- **Objective Evaluation of GSL Recognizer:**
  - ✓ **Evaluation** via **"active"-type exercises** requiring SL production by the learner.
  - ✓ **Numerals**: 3 assignments of six GSL production questions.
  - ✓ **Non-numerals**: 6 assignments of six GSL production questions.
  - ✓ **Fingerspelling**: 6 six-question assignments - **Letters** and **words** not in the training set.
  - ✓ **Volunteers**: **7** A0 - **4** A1 level students, and **1** expert.
    - **Each** performs 3 six-question assignments (one per task, totaling **18 questions**).
  - ✓ "Active"-type **exams** are automatically **graded** by the **system**.
  - ✓ **Results** are **better** than **SI recognition** performance of the **isolated**/**fingerspelling** tasks.
- **Subjective Assessment of the Platform:**
  - ✓ **Participants:** anonymous **subjective experience questionnaire**.
  - ✓ Measures **8 aspects** on the one-to-five **Likert scale**.
  - ✓ In **half questions** most of the users provided the **highest assessment**.



Conclusions

- **Presented the SL-ReDu learning platform GSL recognizer:**
  - ✓ **Isolated** signs and continuously **fingerspelled** sequences.
- **Recognition module:**
  - ✓ **Incorporates** state-of-the-art **deep learning** based visual detection, feature extraction, and classification.
  - ✓ **Operates** in a **SI** fashion in **non-ideal** visual **environments**.
- Designed module **performs very well**, as evidenced by experimental results.
- Yields **very satisfactory objective** and **subjective** user **evaluation** of the SL-ReDu platform.

References

[1] **Potamianos et al.,** "SL-ReDu: Greek sign language recognition for educational applications. Project description and early results," *Proc. PETRA*, **2020**.
[2] **Sapountzaki et al.,** "Educational material organization in a platform for Greek Sign Language self monitoring and assessment," *Proc. EDULEARN*, **2021**.
[3] **Efthimiou et al.,** "The SL-ReDu environment for self-monitoring and objective learner assessment in Greek Sign Language," *Proc. HCII*, **2021**.
[4] **Adaloglou et al.,** "Comprehensive study on sign language recognition methods," *Transactions on Multimedia*, **2022**.

The research work was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "First Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant" (Project Number: 2456).