

# Sign language

Prezentace projektu - Analýza pohybu pro syntézu znakového jazyka

Zdeněk Krňoul

Fakulta aplikovaných věd  
Západočeská univerzita v Plzni

18.10.2021

- 1 Project overview
  - Project aim
- 2 Data-set properties
- 3 Motion capture
  - Recording session
  - Data post-processing
- 4 Integration into the ELG
- 5 Initial results

- **Project name:** Motion-Capture 3D Sign Language Resources
- **Project type:** a) contributes resources, services, tools or data sets to ELG
- **Team:**
  - Miloš Železný, Ph.D. Research Associate Professor at the Department of Cybernetics (DC): *project leader - Machine learning (ML), Artificial intelligence (AI), Computer vision (CV) and Human-machine interaction;*
  - Luděk Müller, Professor, Ph.D.: *Head of AI group at DC, co-author of speech and SL corpora (LDC, ELRA) - ML and AI.*
  - Pavel Jedlička, MSc Ph.D. candidate at UWB: *3D scanning and motion capture analysis;*
  - Zdeněk Krňoul, Ph.D. Assistant Professor at the DC: *3D scanning, ML and AI;*
- **Duration:** from April 2021 to February 2022
- <http://www.kky.zcu.cz/en/research-fields/motion-capture-sign-language>

## Project Aim:

- Create a Motion Capture data-set for Czech sign language (CSE)
- Data-set primary purpose: sign language synthesis and machine translation



<https://simax.media/>



## Weather forecast

- 36 individual forecasts (one forecast  $\sim$  30sec continuous speech)
- Vocabulary diversity (3 forecasts per month)
- Large sample of numbers (temperatures)
- Limited vocabulary - multiple instances of same sign in different context (frequent signs more than 20 repetitions)

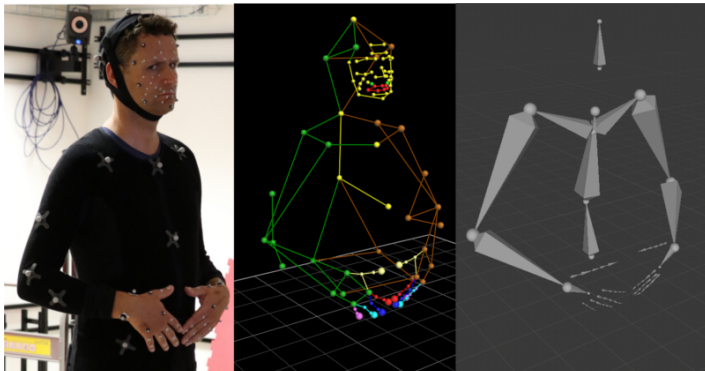
## Zoo tour

- 20 different animals
- Structured description - repetition of similar signs (biotope, food, lifespan, ...)
- Large sample of letters (Latin names - finger-spelling)
- (extensibility)

Summary: data-set contains multiple instances of same signs in different grammar context

# Motion capture principle: VICON System

- Technology provides recording with the marker position error of less than 1 mm (in 3D space)
- Subject is equipped with **retro-reflexive markers**
- Subject movements are recorded by multiple cameras simultaneously
- 3D positions of markers and their trajectory are calculated and reconstructed in 3D
- Solving **skeleton structure** according to 3D avatar model



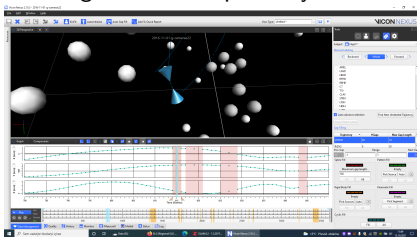
## Motion capture recording session:

- 1 Camera placement and calibration (setup)
- 2 Subject preparation, marker placement, and calibration of range of movements (ROM)
- 3 **Hand-shape, body, facial movements** recorded separately (different camera setups)
- 4 Movement recording - short takes (longer takes than 1 minute are recorded in parts)
  - Recording day - max 60 min of continuous signing

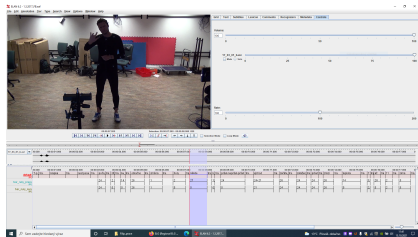




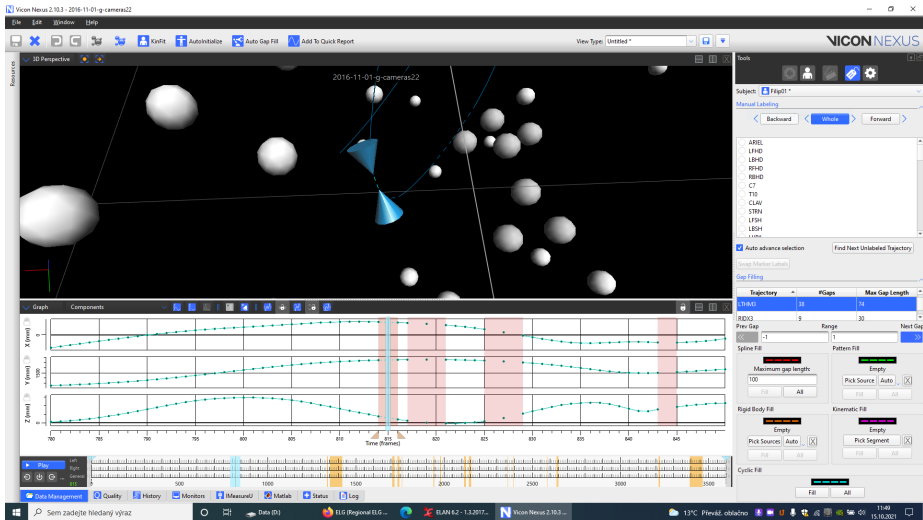
- Cleaning the data acquired by the Motion Capture system (noise, occlusions, ...)



- Data annotations (from reference camera)



- Missing data caused by body parts occlusions, perspective, marker alignment, etc.
- Manually filled (assisted interpolation using patterns, rigid parts of body, ...)





- **Data:**
  - Cleared motion data (continuous speech + dictionary)
  - Data for SL handshapes and facial cues
- **Tool** for searching in the data-set
  - Input: query as short-motion action
  - Output: all relevant results from the data-set

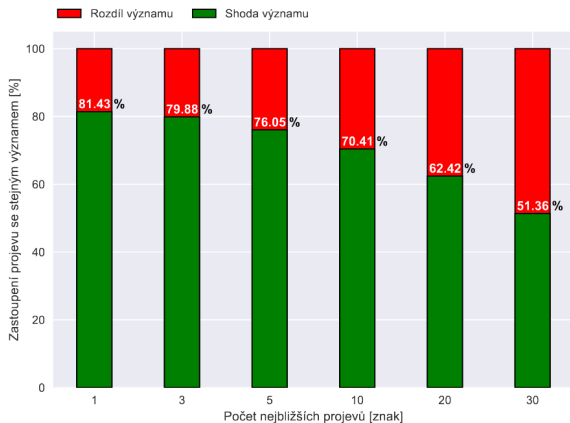
## Data usability:

- SL synthesis
- SL analysis (grammar, machine translation)
- Signer differences analysis
- SL Animation re-targeting
- ...

- ELG project: 20 hrs
- 5 signers selected (2 more in process of selection)
- Weather Forecast: 2 hrs / 4,5 (from 3 SL speakers)
- Zoo: 2 hrs / ~ 21 (from 1 SL speaker)

# Initial results: Searching glosses

- Measuring similarity between two temporal sequences
- Experiment for full skeleton data
- Baseline algorithm: dynamic time warping (DTW)
- Accuracy  $\sim 70\%$  for 10 best matches



Thank you for your attention