

# **Real-Time Multilingual Sign Language Processing**

Amit Moryossef

Ph.D. Thesis

# Transparency

## **Academic Affiliations:**

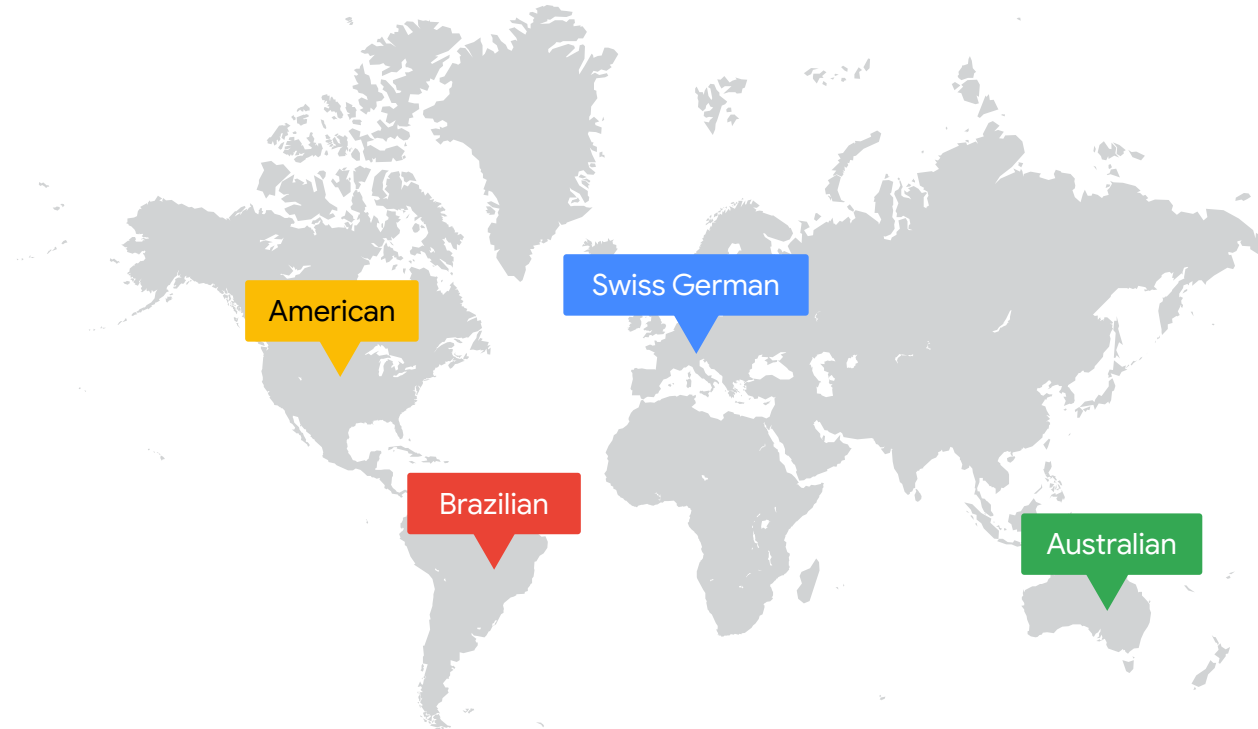
- Bar-Ilan University (Israel) - Graduating Ph.D. Student
- University of Zurich (Switzerland) - Postdoctoral Researcher (SIGMA Project, DSI)

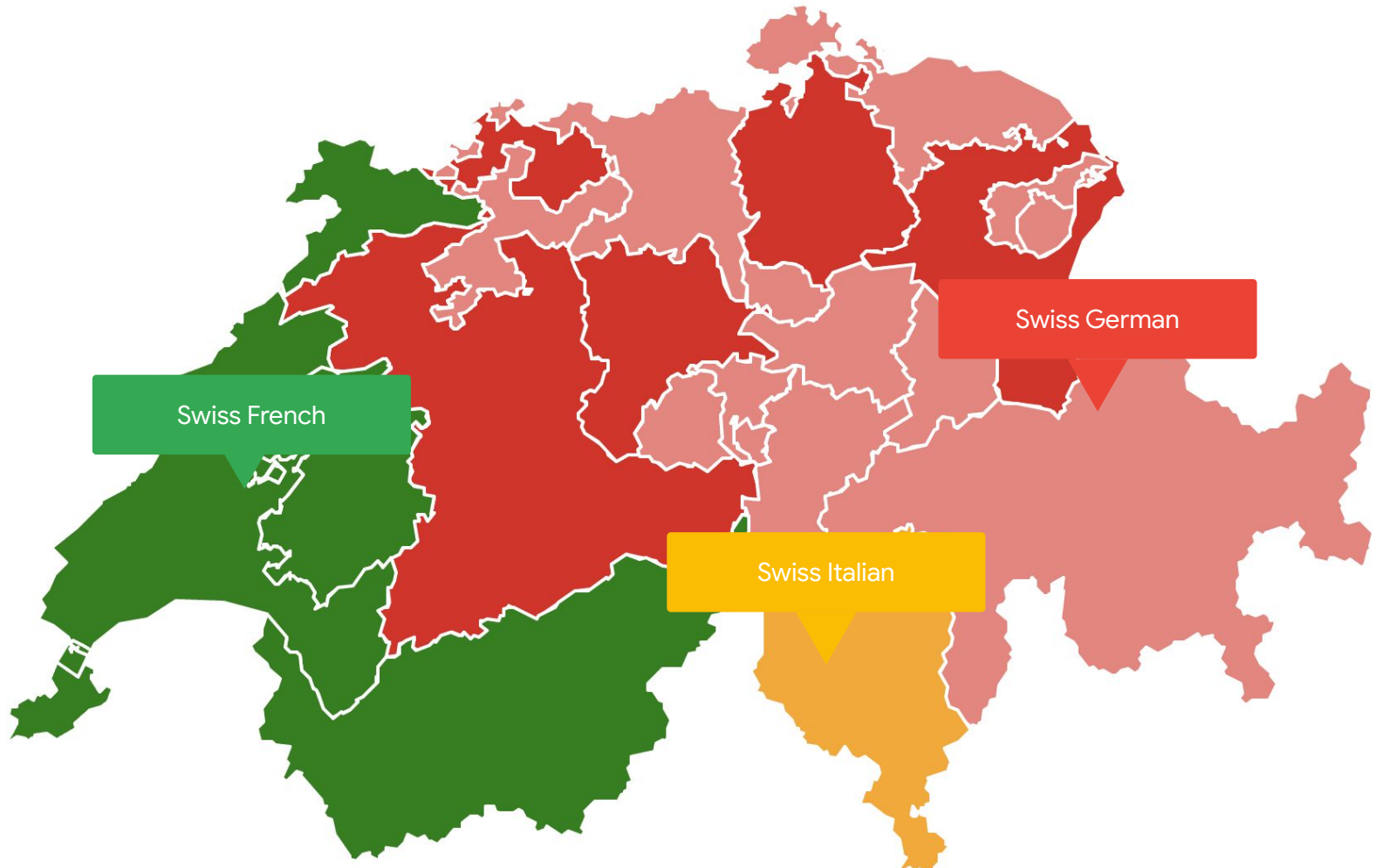
## **Industry:**

- Owner and only employee of *sign.mt ltd*, looking for funding

# Signed Languages

Signed languages are the primary means of communication for many deaf and hard of hearing individuals.





DSGS: [Zürich](#), [Bern](#), [Basel](#), [Lucerne](#), and [St. Gallen](#), as well as in [Liechtenstein](#).

# (Goal?) of Existing Sign Language Works



How to sign "hello" in asl?



[All](#)

[Videos](#)

[Images](#)

[Books](#)

[News](#)

[More](#)


[Settings](#)

[Tools](#)

About 623,000,000 results (0.46 seconds)

English – detected ↔ American Sign Language (ASL)

Hello ×



Speaker icon | Microphone icon

[Open in Google Translate](#)

[Feedback](#)

**tiny**

**the tiny corp** 

@\_\_tinygrad\_\_



The hard part of programming is finding the right way to factorize the problem. The rest is fill in the blank.

12:19 PM · Feb 10, 2024

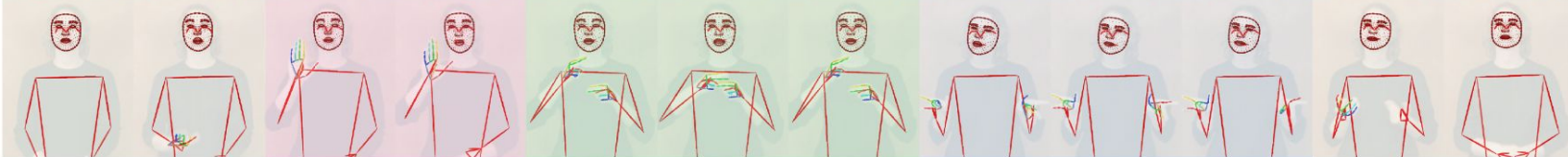
# Representations of Signed Languages

What is your name?

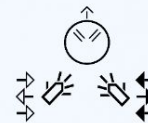
Video Stream



Pose Stream



SignWriting



HamNoSys

$\emptyset \wedge \emptyset \uparrow$

$[\text{d}_1 \text{e}_0 \text{z} \text{d}_1 \text{u}_0] \text{u}_2 \text{)(} \downarrow \text{x} +$

$\cdot \bar{\emptyset} \setminus \emptyset \text{u}_0 \text{u}_1 \cdot (\uparrow \circ) +$

ASL Gloss

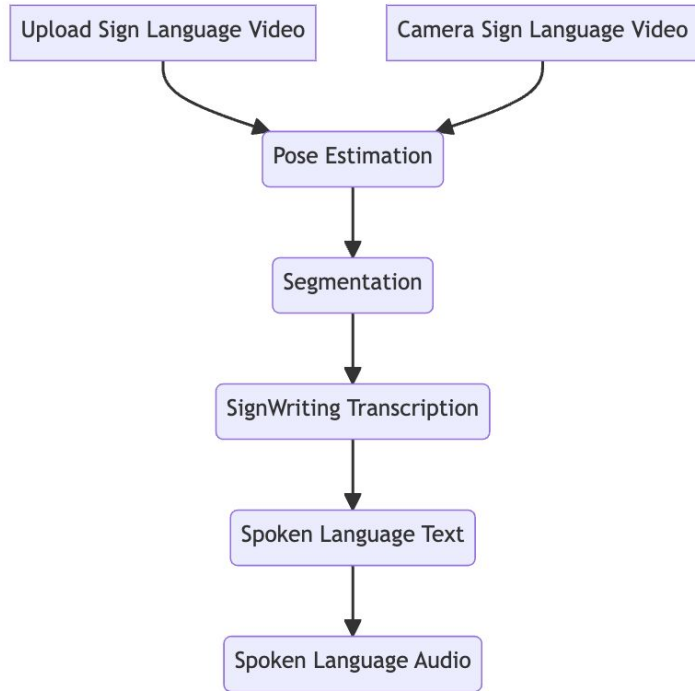
YOUR

NAME

WHAT

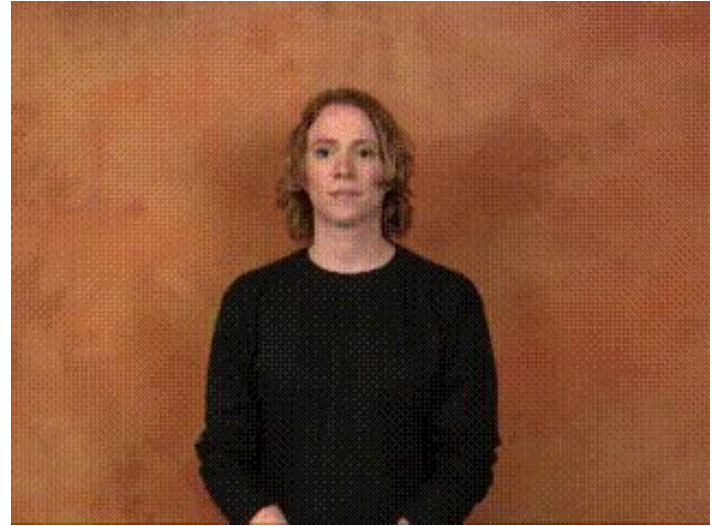
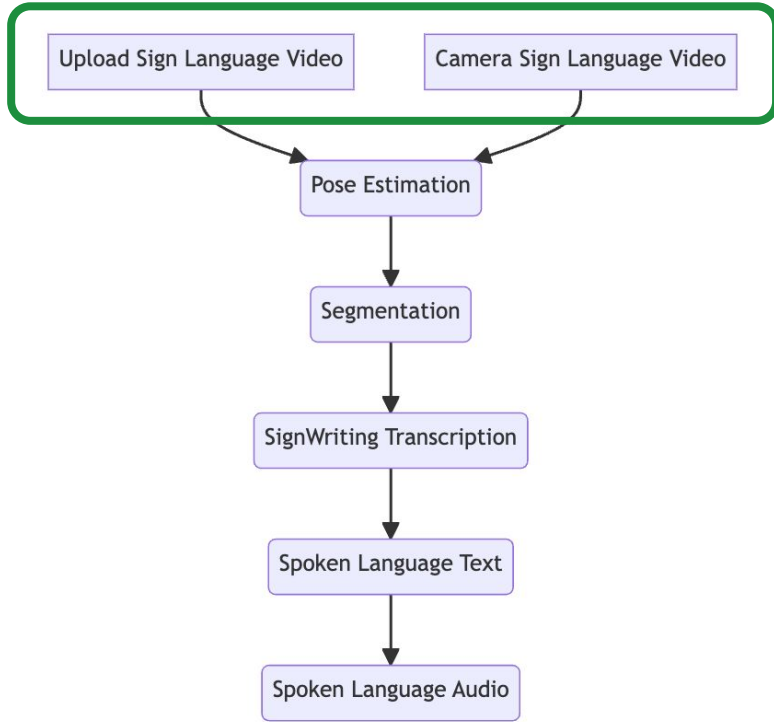
107 FRAMES

# The Signed-to-Spoken Translation Pipeline

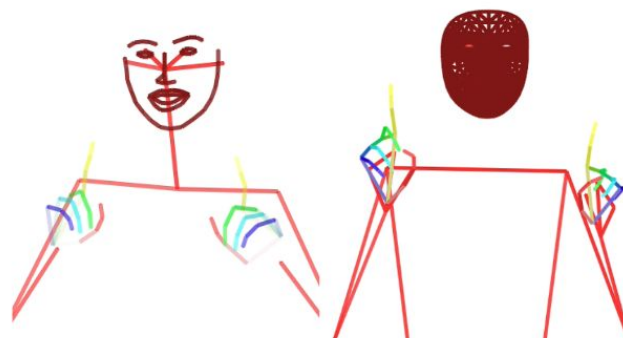
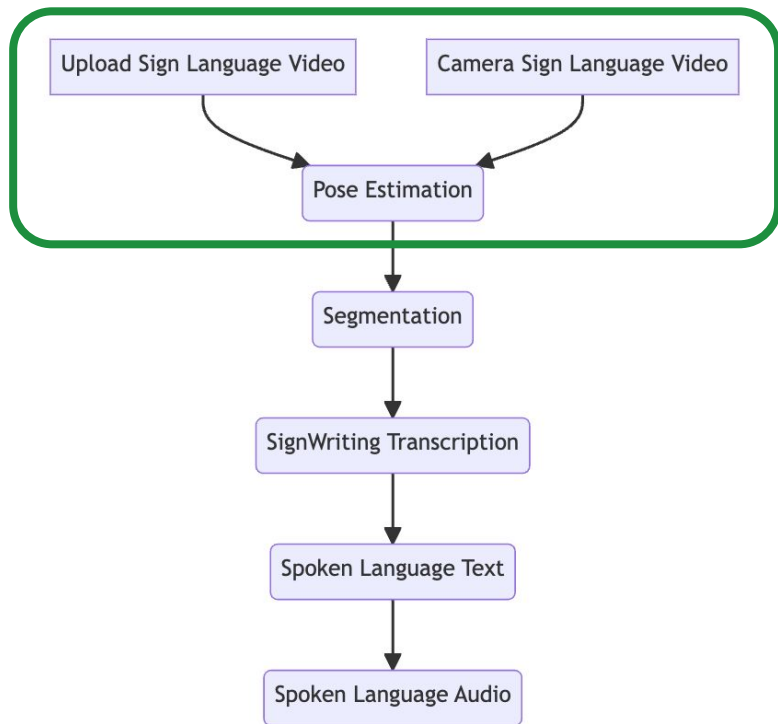




# The Signed-to-Spoken Translation Pipeline

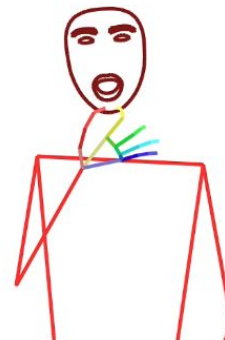


# The Signed-to-Spoken Translation Pipeline

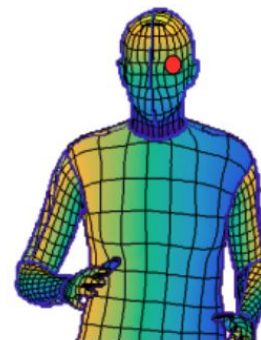


(a) OpenPose

(b) Mediapipe



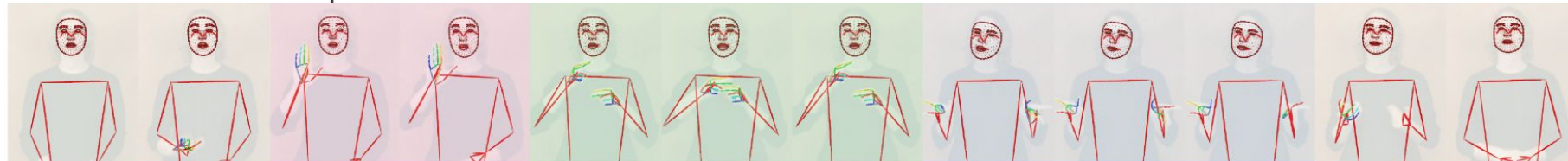
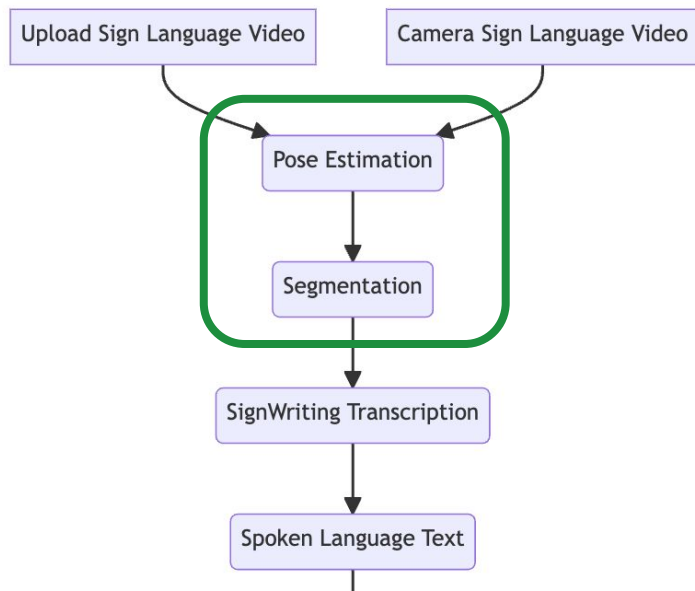
(c) 2b w/ Postprocessing



(d) DensePose

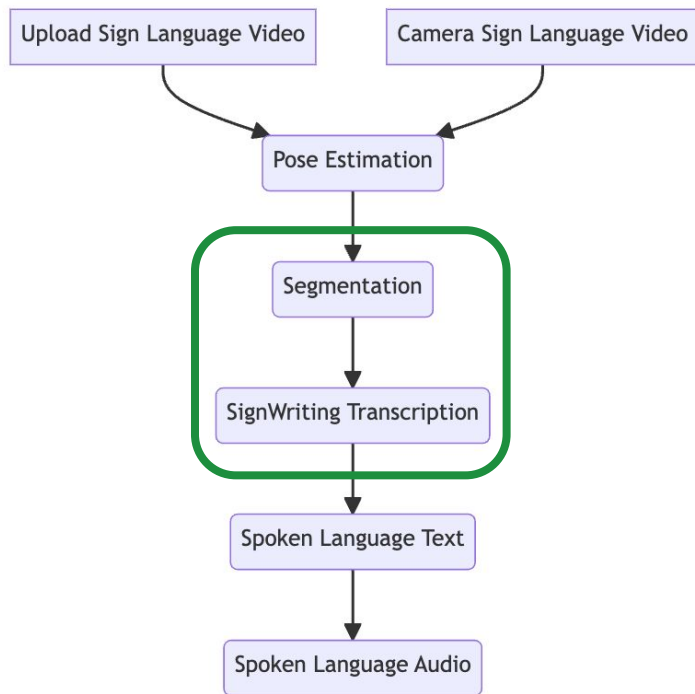
- Ivan Grishchenko and Valentin Bazarevsky. 2020. Mediapipe holistic.
- Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. 2019. OpenPose: Realtime multi-person 2D pose estimation using part affinity fields.

# The Signed-to-Spoken Translation Pipeline



- Amit Moryossef, Zifan Jiang, Mathias Müller, Sarah Ebling, and Yoav Goldberg. Linguistically motivated sign language segmentation. <https://github.com/sign-language-processing/segmentation>

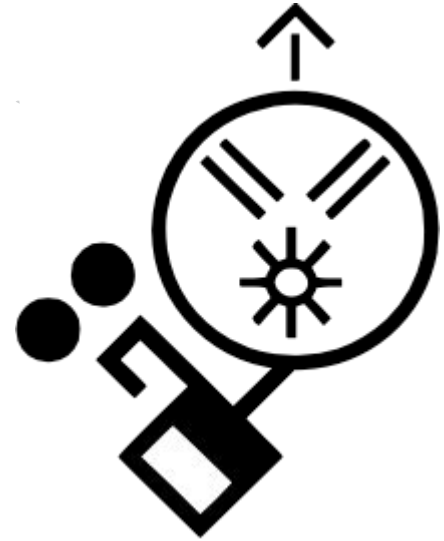
# The Signed-to-Spoken Translation Pipeline



- Work in progress <https://github.com/sign-language-processing/signwriting-transcription>

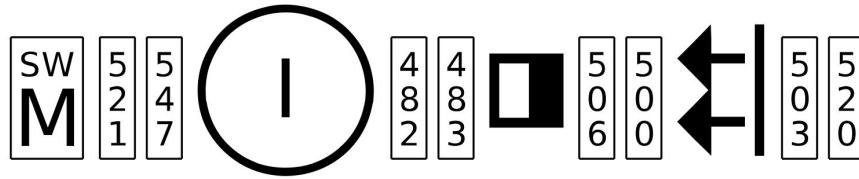
# SignWriting Construction

- **FACE**
  - Eyebrows: **STRAIGHT DOWN**
  - Mouth: **MOUTH OPEN WRINKLED**
  - Movement: **FLOORPLANE SINGLE STRAIGHT SMALL**
- **HAND**
  - Shape: **HAND-FIST INDEX THUMB SIDE INDEX BENT**
  - Handedness: **RIGHT**
  - Plane: **WALL**
  - Facing: **SIDE**
  - Rotation:  $\frac{1}{8}$  (45 degrees anti-clockwise)
  - Contact: **CHIN**
  - Movement: **SQUEEZE LARGE MULTIPLE**

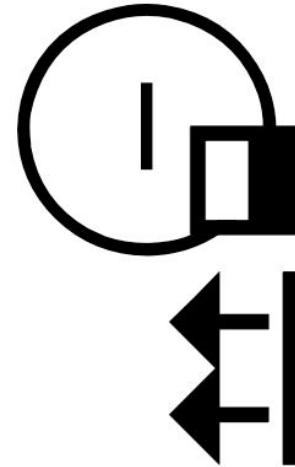


# SignWriting Representation

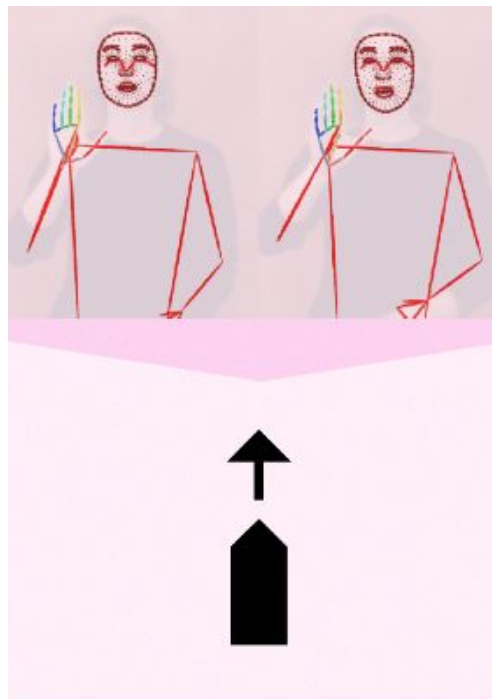
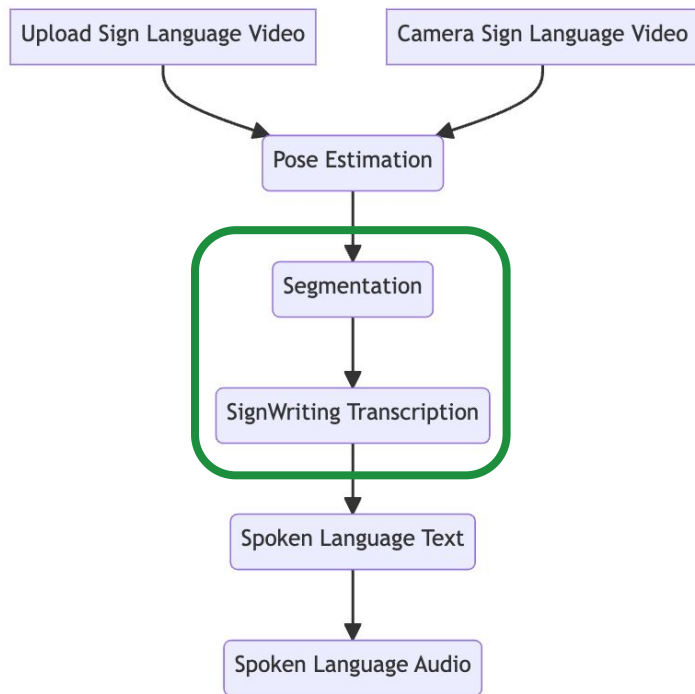
SignWriting1D



SignWriting2D

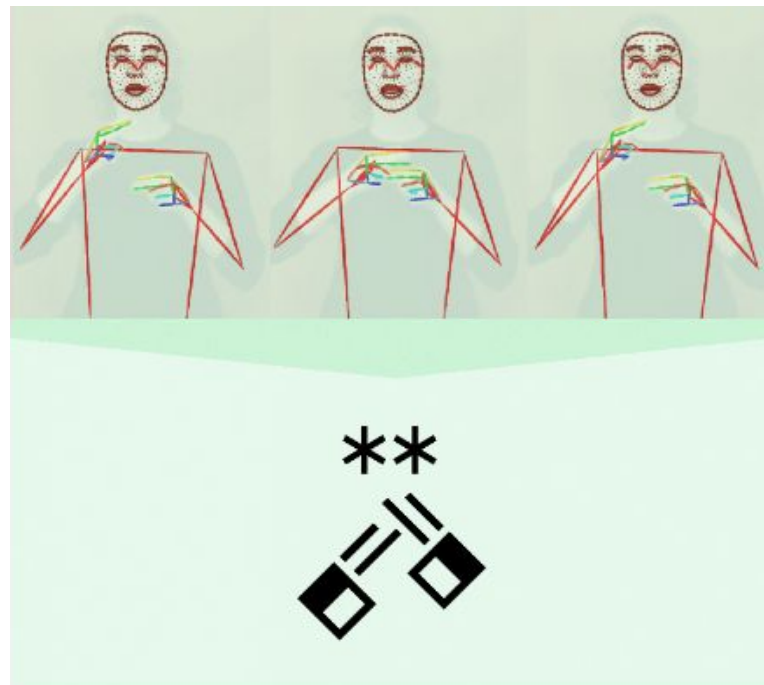
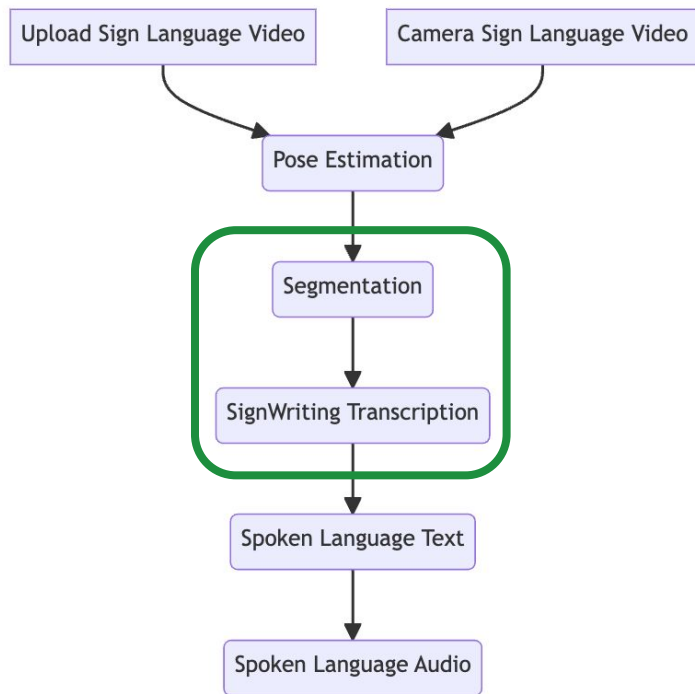


# The Signed-to-Spoken Translation Pipeline



- Work in progress <https://github.com/sign-language-processing/signwriting-transcription>

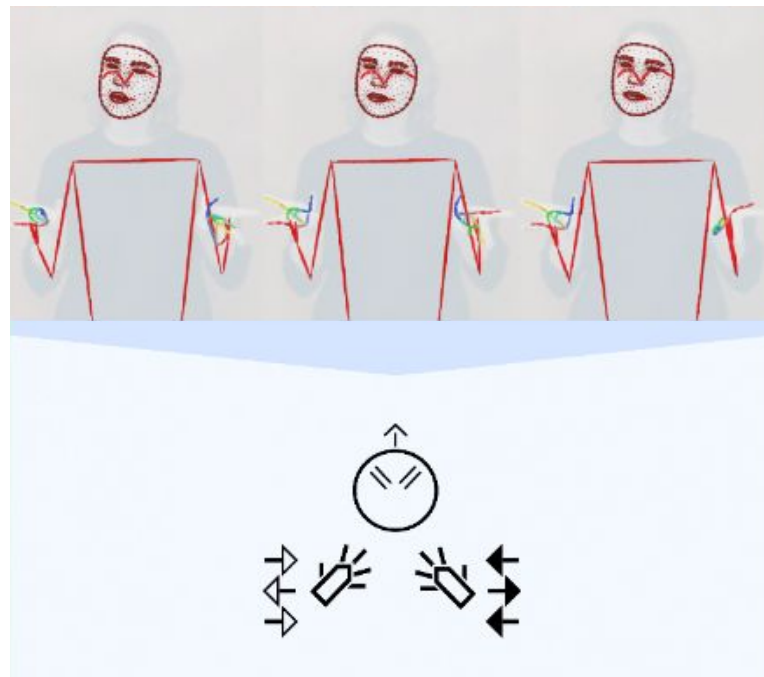
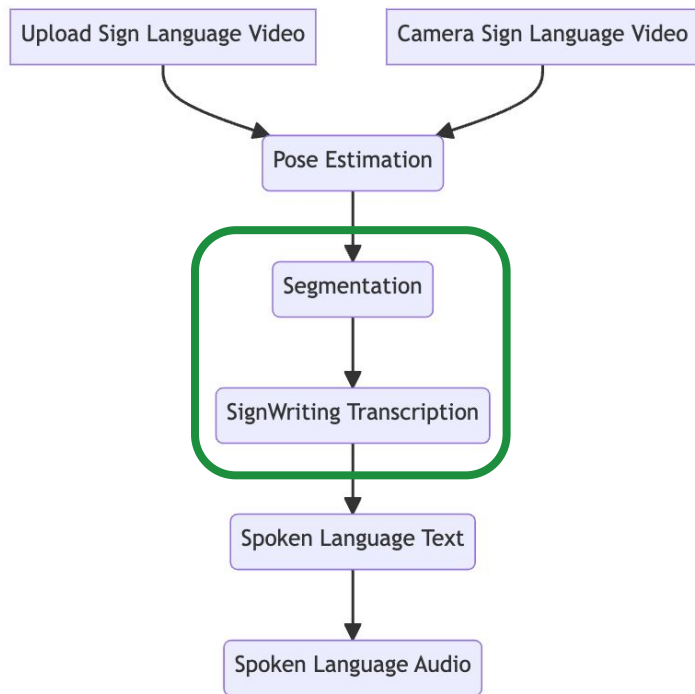
# The Signed-to-Spoken Translation Pipeline



- Work in progress <https://github.com/sign-language-processing/signwriting-transcription>

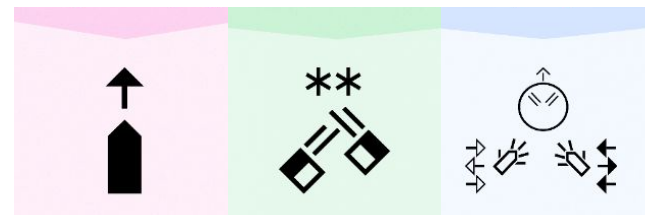
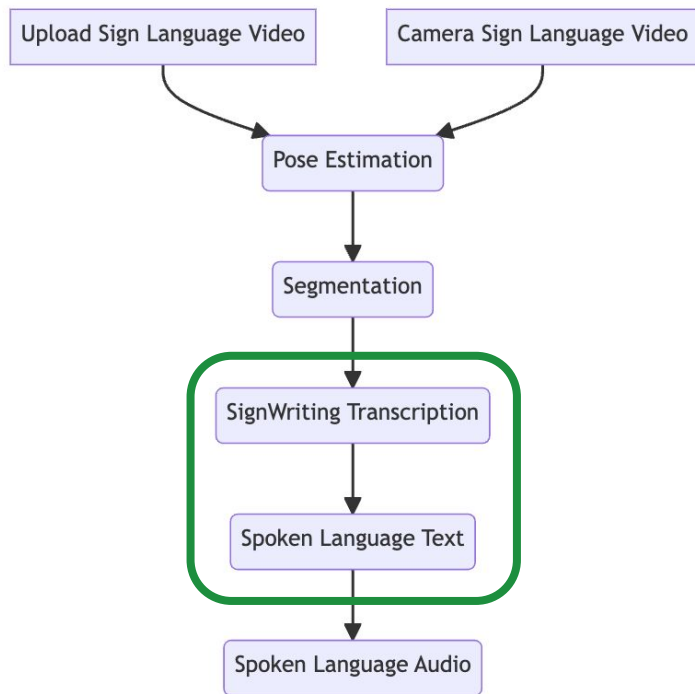


# The Signed-to-Spoken Translation Pipeline



- Work in progress <https://github.com/sign-language-processing/signwriting-transcription>

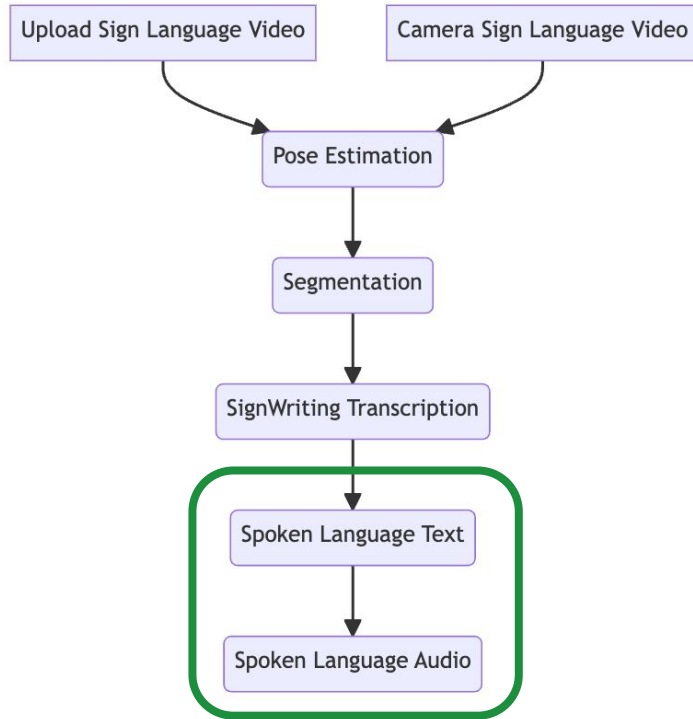
# The Signed-to-Spoken Translation Pipeline



What is your name?

- Jiang, Z., Moryossef, A., Müller, M., & Ebling, S. (2022). Machine Translation between Spoken Languages and Signed Languages Represented in SignWriting.
- Moryossef, A., & Jiang, Z. (2023). SignBank+: Preparing a Multilingual Sign Language Dataset for Machine Translation Using Large Language Models.

# The Signed-to-Spoken Translation Pipeline



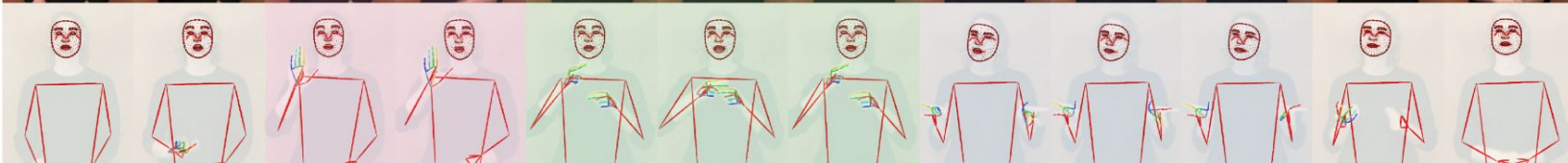
Native speech synthesis

# The Signed-to-Spoken Pipeline in Theory

Video Stream

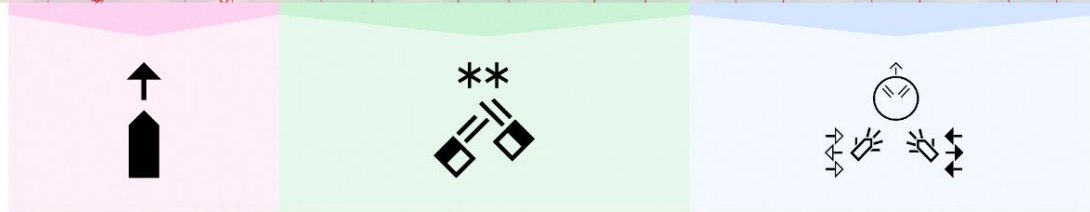


Pose Stream



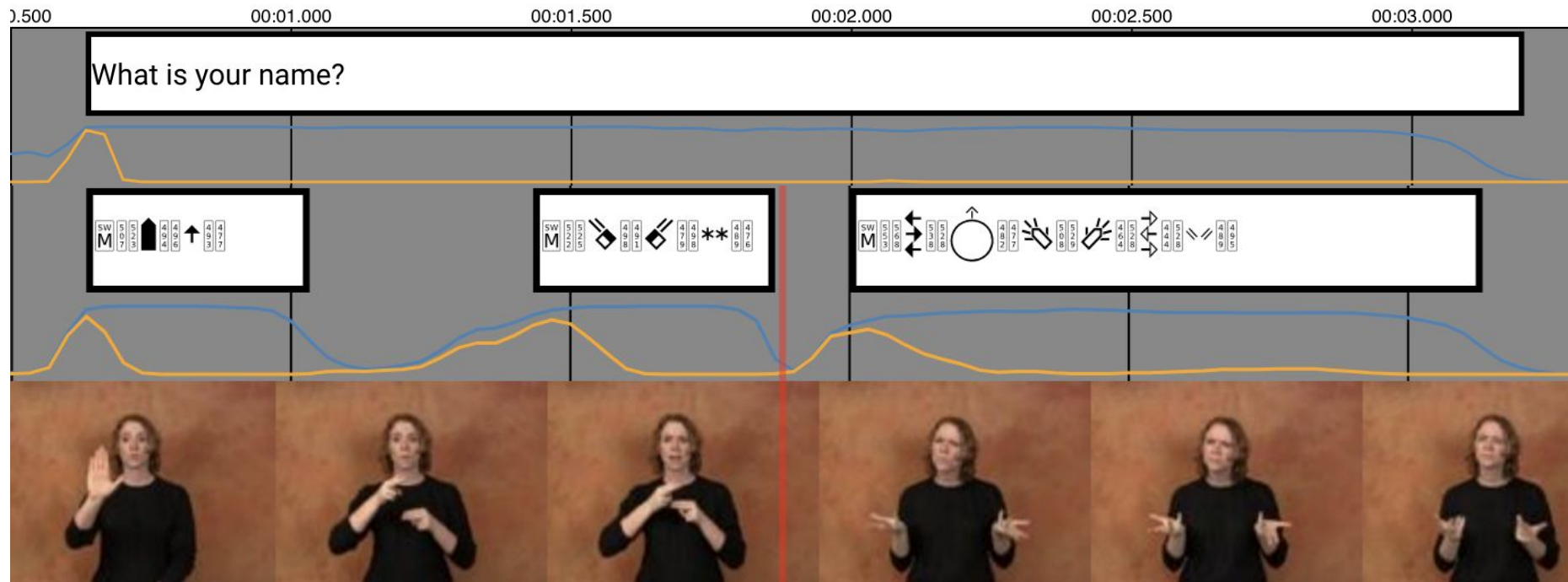
107 FRAMES

SignWriting

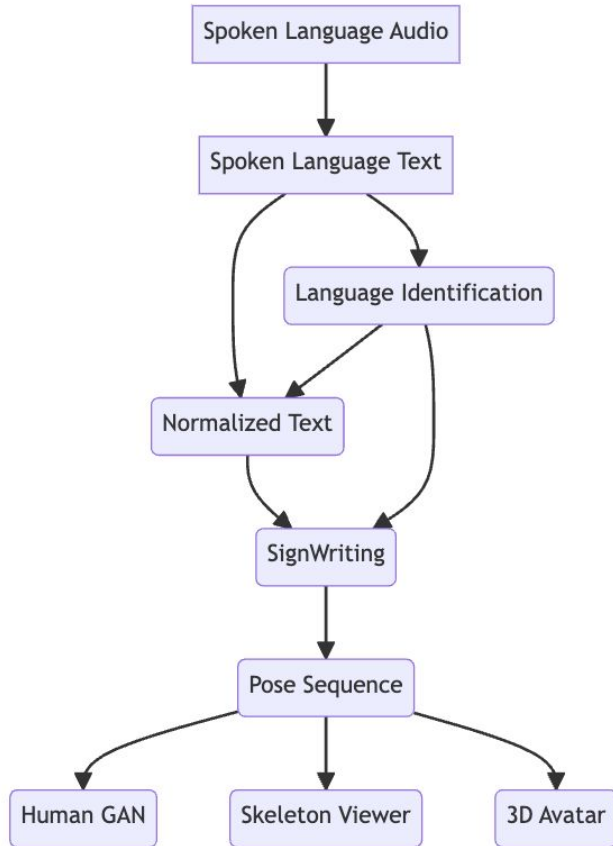


What is your name?

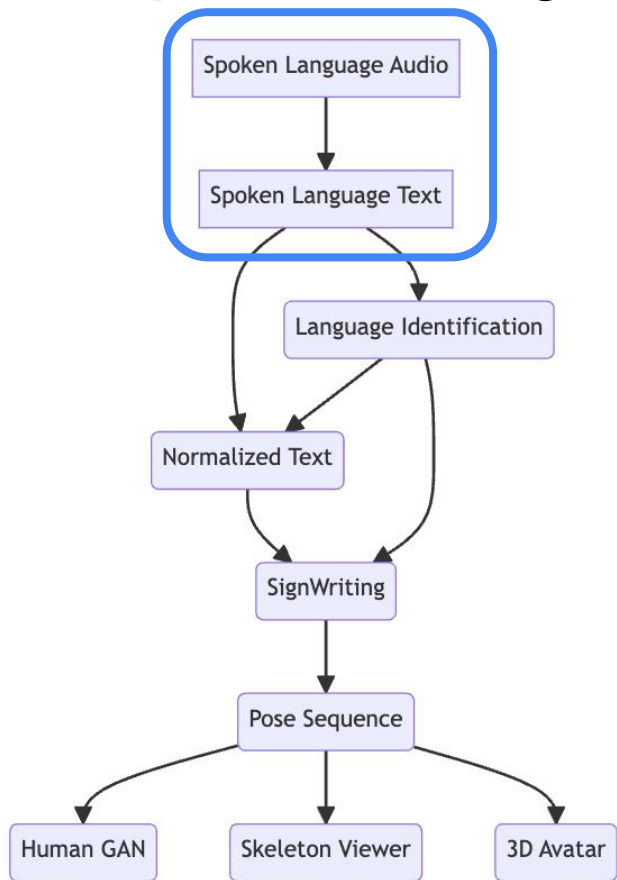
# The Signed-to-Spoken Pipeline in Practice!



# The Spoken-to-Signed Translation Pipeline

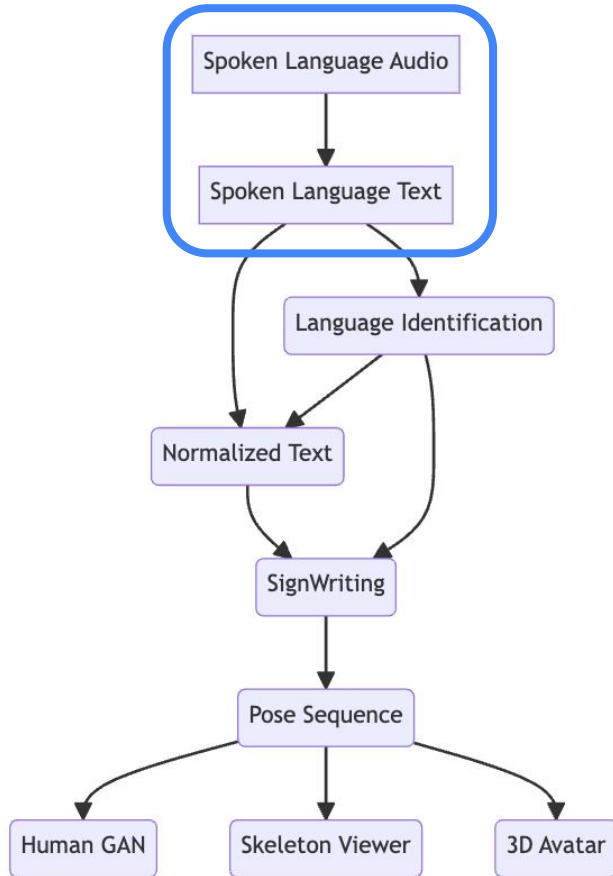


# The Spoken-to-Signed Translation Pipeline



Native speech recognition

# Browser Speech Recognition



**Implemented**

Recommended. Fast and simple.

**Supported browsers:**

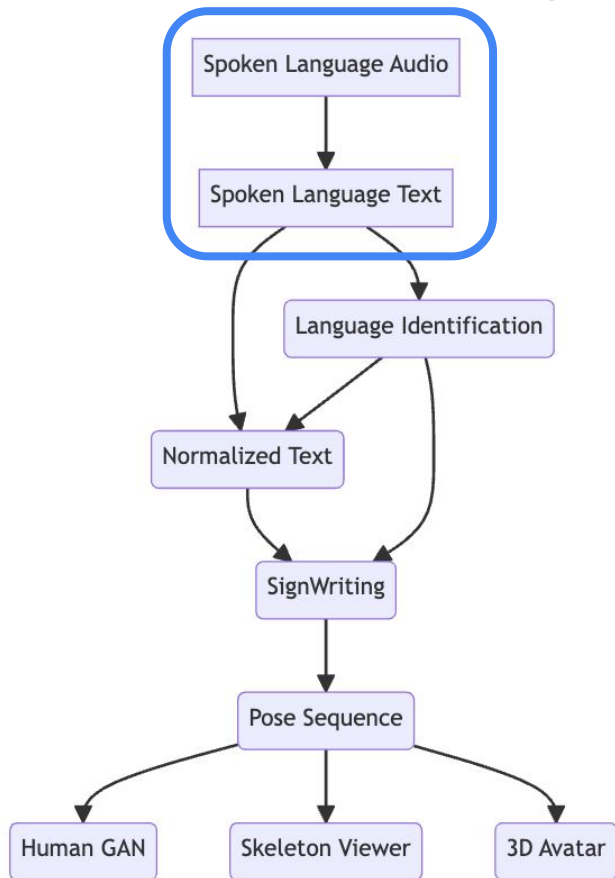
<https://caniuse.com/?search=SpeechRecognition>

**Supported languages:**

Device dependent



# Custom Model (e.g. Whisper)



**Not Implemented**

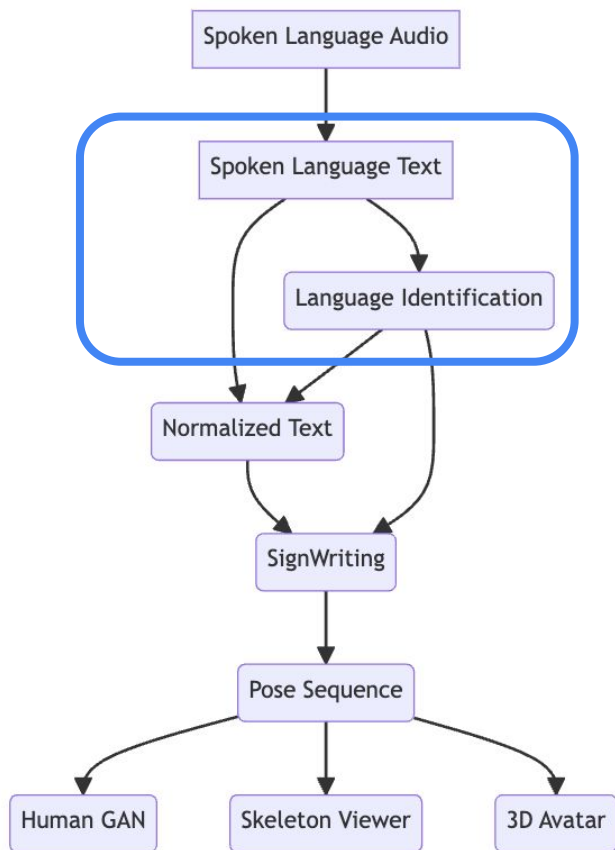
Not generally recommended.  
Requires loading and is slower.

Should be used when the browser does not support speech recognition (e.g. firefox) or when the language is not natively supported.

**Supported browsers:** all

**Supported languages:**  
TODO (based on model)

# The Spoken-to-Signed Translation Pipeline



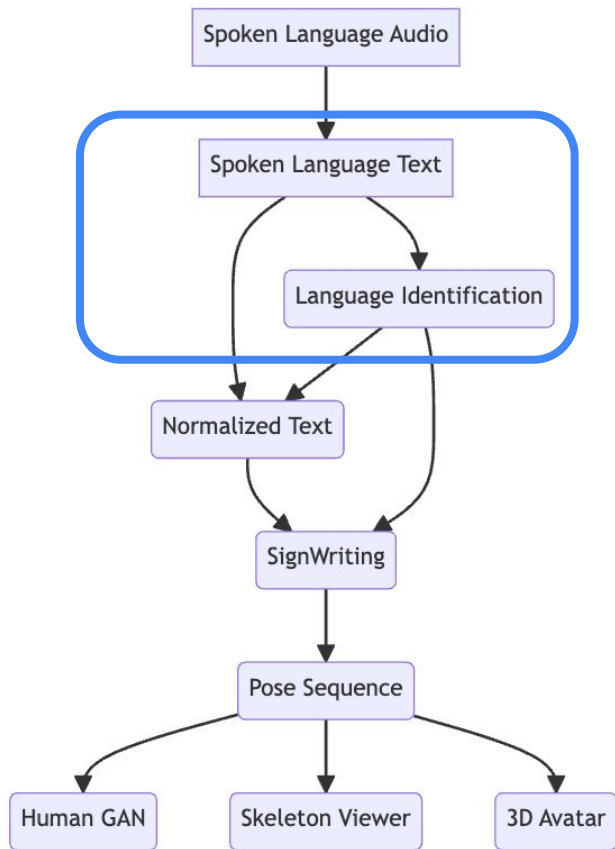
Off-the-shelf Language Identification

DETECT LANGUAGE ENGLISH GERMAN FRENCH

Kleine Kinder essen Pizza

✦ Translate from: German

# Google's Compact Language Detector 3



Implemented

<https://github.com/kwonoj/cld3-asm>

Model size: 1.1Mb

**Supported languages:**

<https://github.com/google/cld3#supported-languages>

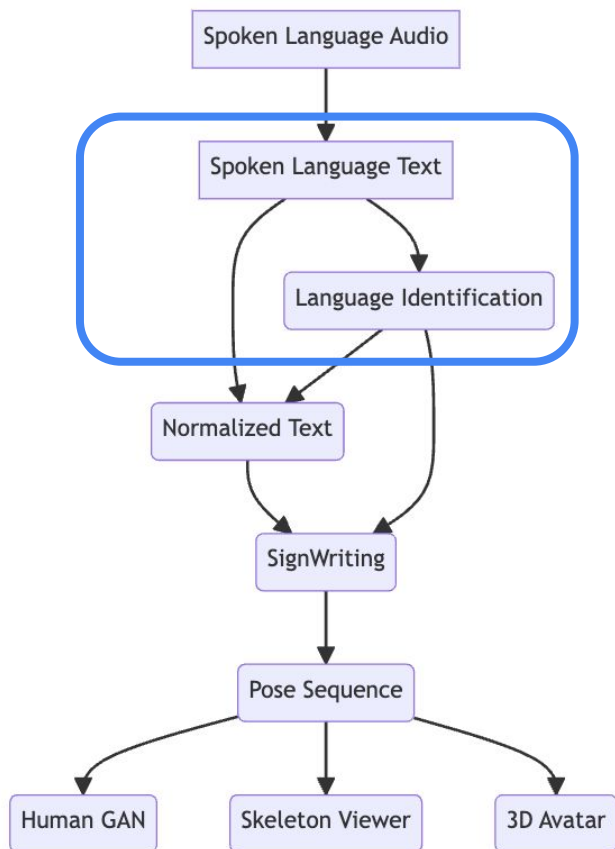
**Issues:** struggles with short texts

<https://github.com/google/cld3/issues/76#issuecomment-1625233427>

# MediaPipe Language Detector



MediaPipe



Implemented

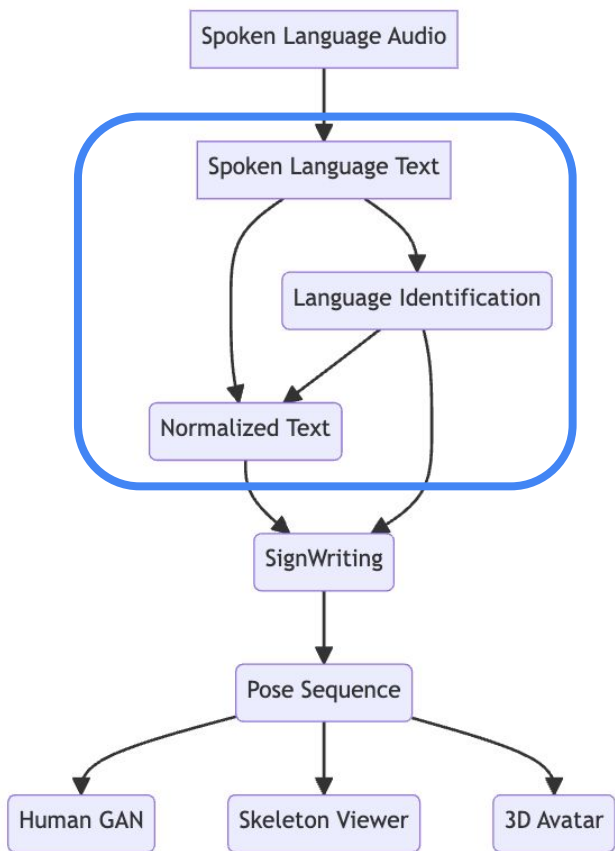
Recommended since cld3 has issues with short texts.

[https://developers.google.com/mediapipe/solutions/text/language\\_detector](https://developers.google.com/mediapipe/solutions/text/language_detector)

Model size: 315kB

Supported languages: [110](#)

# The Spoken-to-Signed Translation Pipeline



## Input (Erroneous)

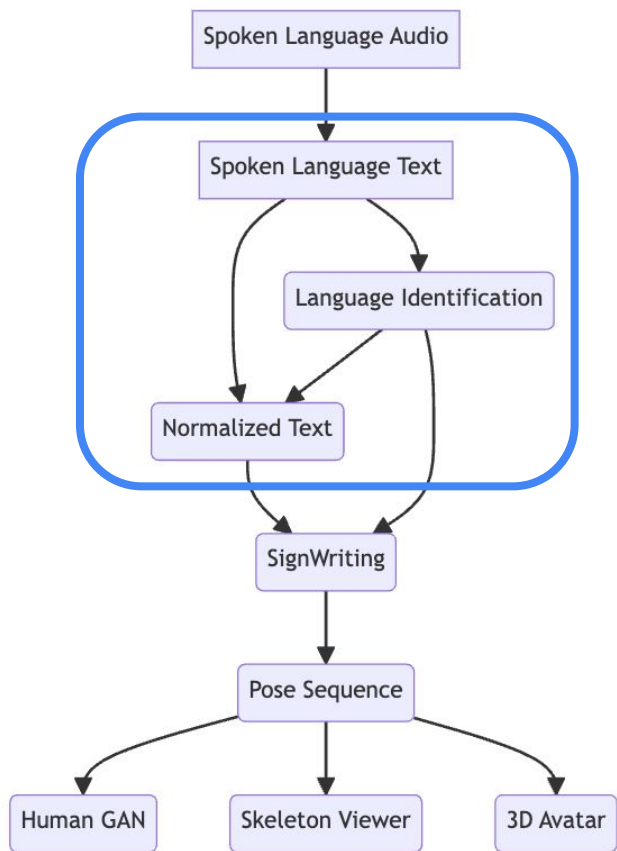
A important part of my life have been a people that stood by me.

## Output (Corrected)

An important part of my life has been the people who stood by me.

Large Language Models

# The Spoken-to-Signed Translation Pipeline



A important part of my life have been  
a people that stood by me

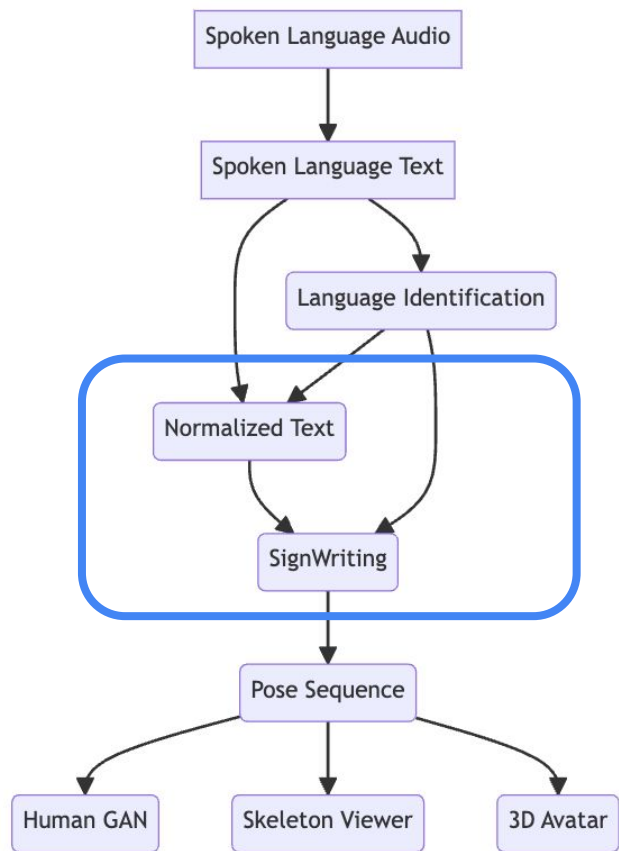
✧ Did you mean: *An important part of my life has been  
the people who stood by me.*

kleine kinder essen pizza

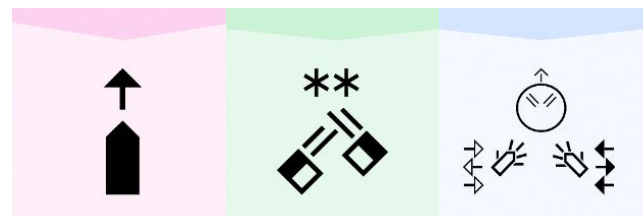
✧ Did you mean: *Kleine Kinder essen Pizza.*

gpt-3.5-turbo

# The Spoken-to-Signed Translation Pipeline

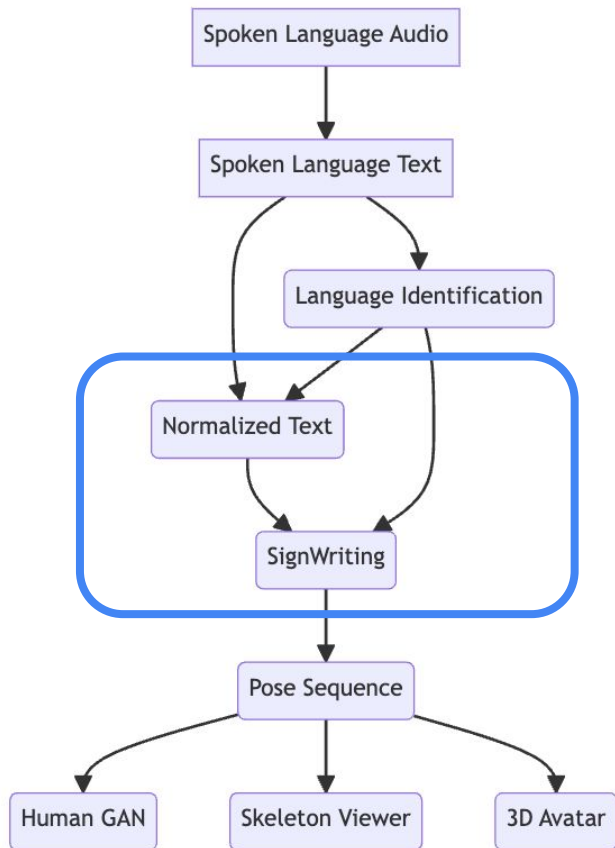


What is your name?



- Jiang, Z., Moryossef, A., Müller, M., & Ebling, S. (2022). Machine Translation between Spoken Languages and Signed Languages Represented in SignWriting.

# The Spoken-to-Signed Translation Pipeline



ENGLISH

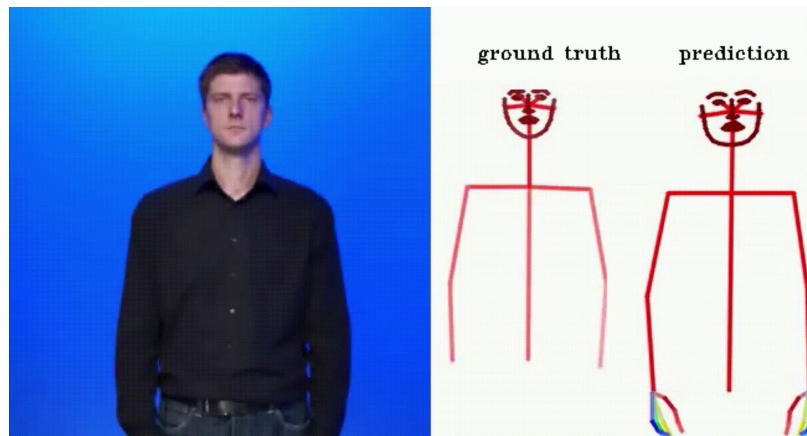
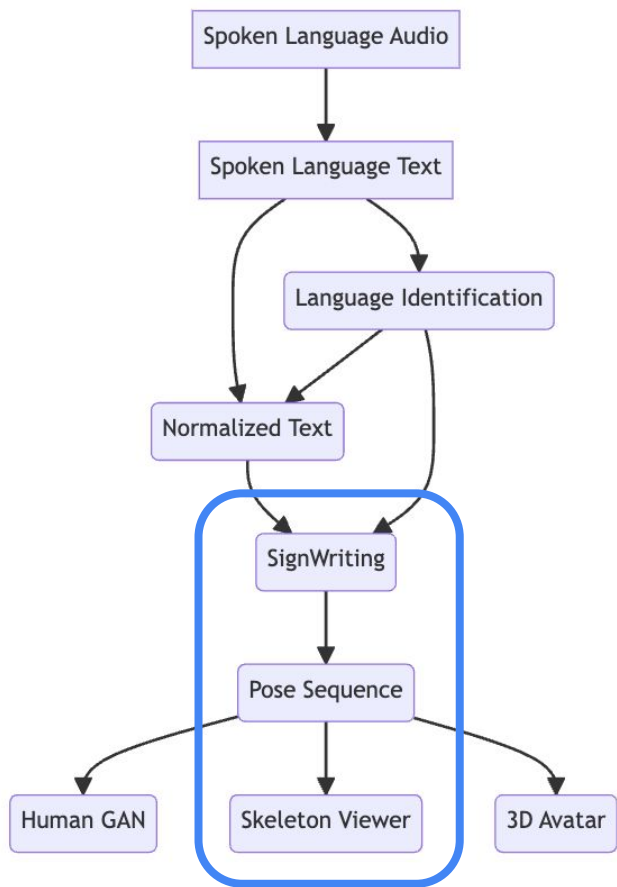
Hello, my name is John.

23 / 500

The image shows a digital interface for sign language translation. At the top, the word 'ENGLISH' is displayed. Below it, the text 'Hello, my name is John.' is shown. To the right of the text is a vertical column of sign language icons representing the sentence. At the bottom left, there are microphone and speaker icons. At the bottom right, the text '23 / 500' is visible.



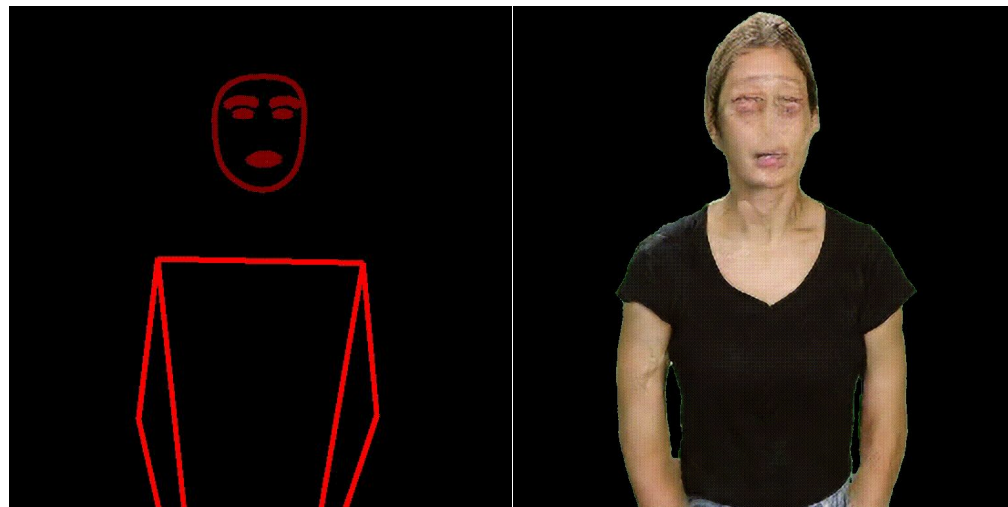
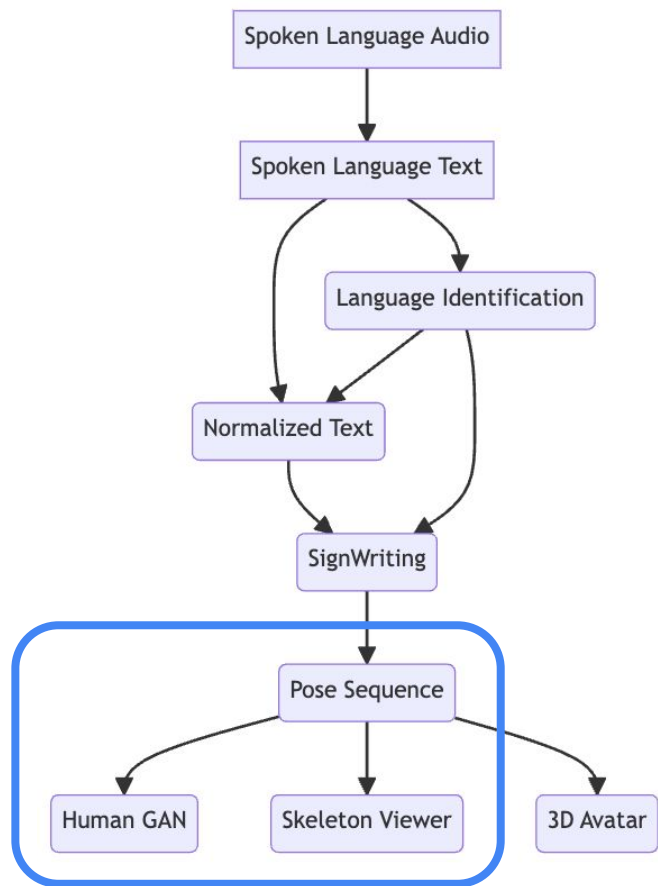
# The Spoken-to-Signed Translation Pipeline



## Motion Diffusion Model

- Arkushin, R. S., Moryossef, A., & Fried, O. (2023). Ham2Pose: Animating Sign Language Notation Into Pose Sequences.

# The Spoken-to-Signed Translation Pipeline



- Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks.
- YouTube video [https://www.youtube.com/watch?v=JpM2\\_IzqePk](https://www.youtube.com/watch?v=JpM2_IzqePk)

# *pose-to-video*: Render pose sequences as photorealistic videos.

## Implementations

---

This repository includes multiple implementations.

### Conditional Implementation

- [pix\\_to\\_pix](#) - Pix2Pix model for video generation
- [controlnet](#) - ControlNet model for video generation

### Unconditional Implementation (Controlled)

- [stylegan3](#) - StyleGAN3 model for video generation
- [mixamo](#) - Mixamo 3D avatar

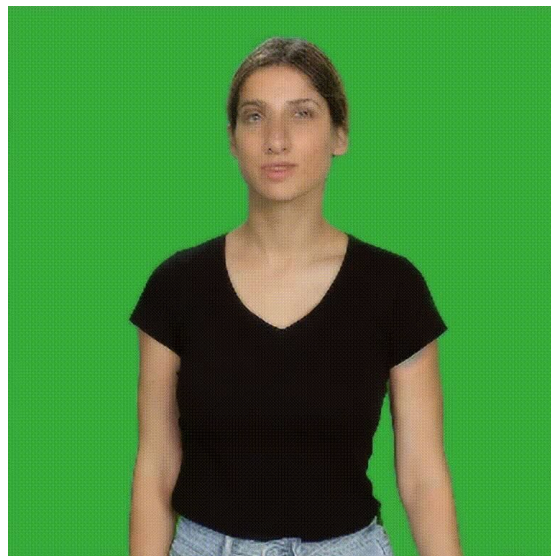
### Upscalers

- [simple-upscaler](#) - Upscales 256x256 frames to 768x768

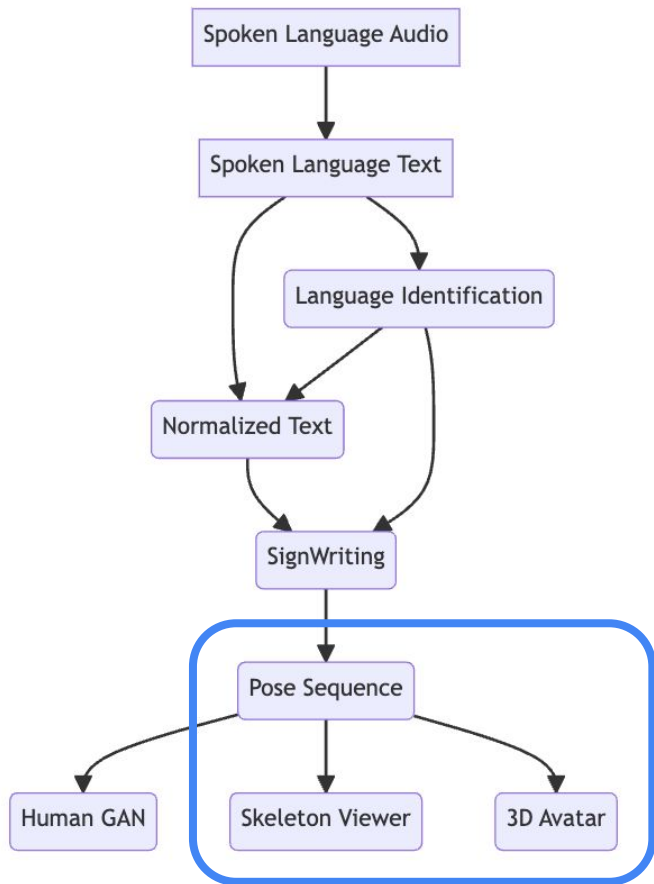
## Datasets

---

- [BIU-MG](#) - Bar-Ilan University: Maayan Gazuli
- [SHHQ](#) - high-quality full-body human images

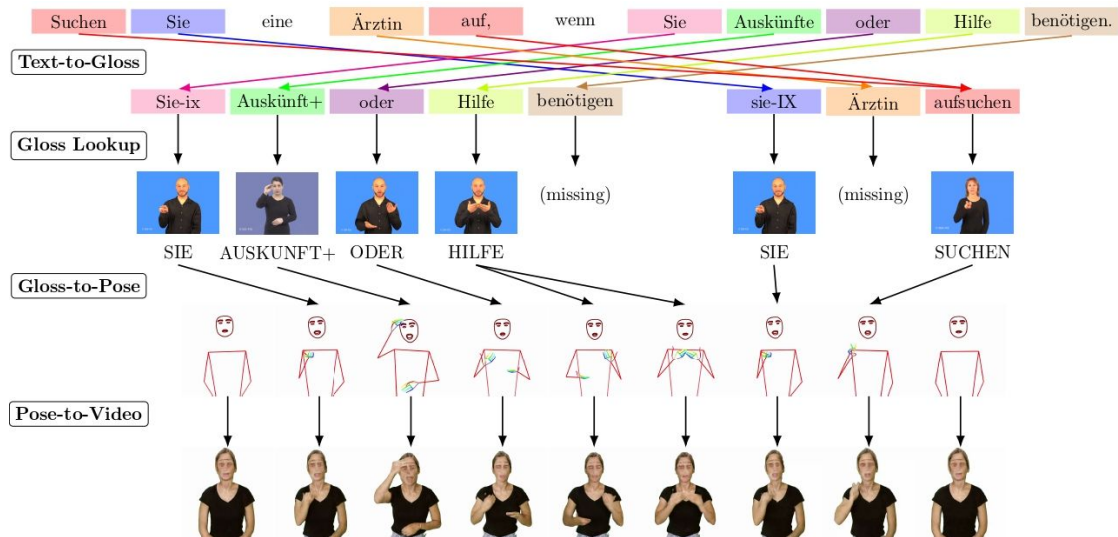
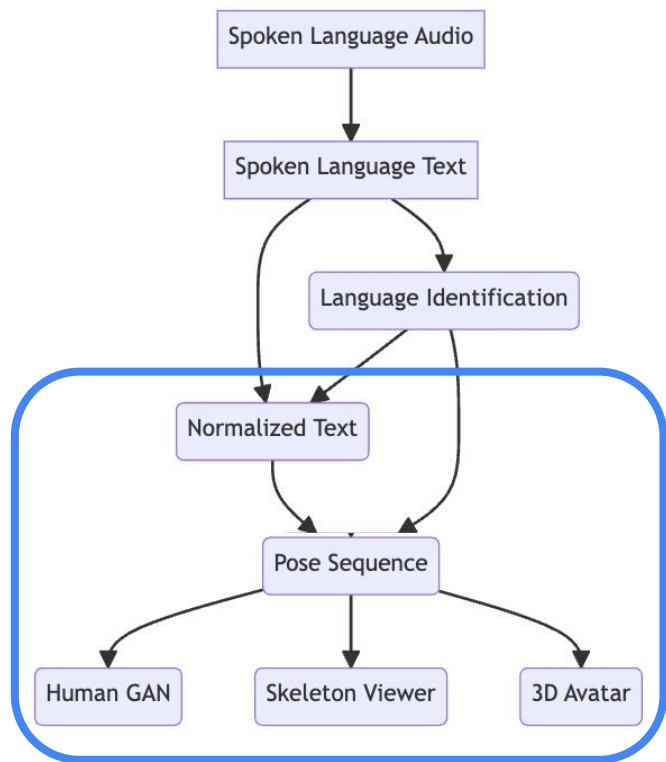


# The Spoken-to-Signed Translation Pipeline



- YouTube video [https://www.youtube.com/watch?v=TyJuU9\\_GOaw](https://www.youtube.com/watch?v=TyJuU9_GOaw)

# The (baseline) Spoken-to-Signed Translation Pipeline

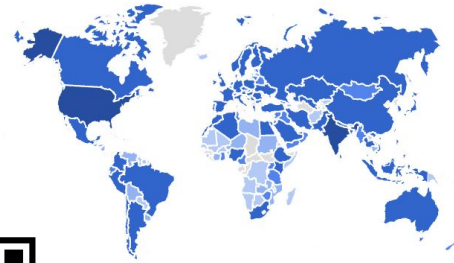
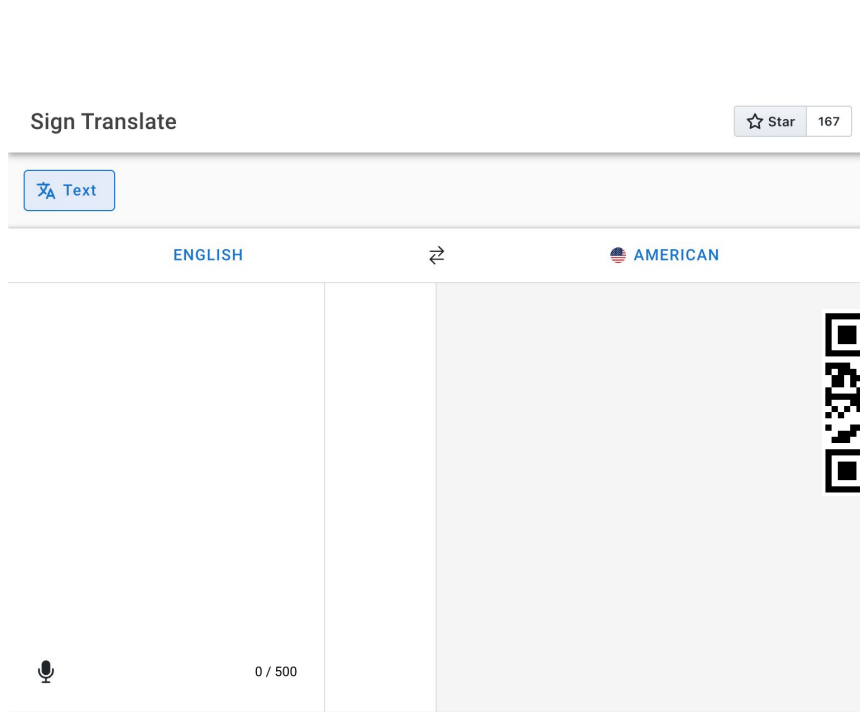


- Moryossef, A., Müller, M., Göhring, A., Jiang, Z., Goldberg, Y., & Ebling, S. (2023). An Open-Source Gloss-Based Baseline for Spoken to Signed Language Translation. <https://github.com/ZurichNLP/spoken-to-signed-translation>

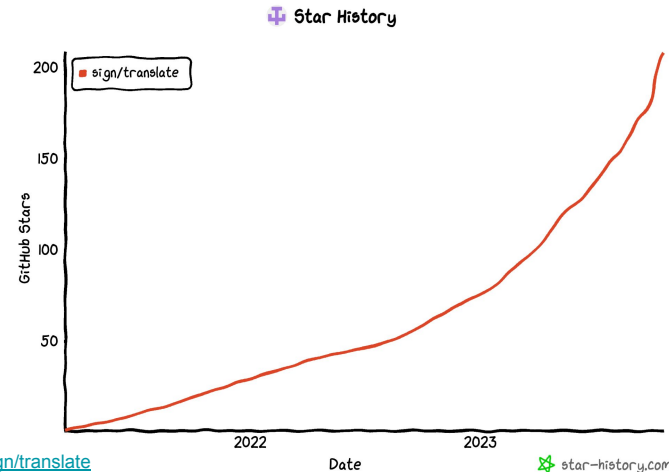
# The (baseline) Spoken-to-Signed Translation Pipeline in Action!

The image shows a Google search interface. The search bar contains the text: "Wie gebärdet man "kleine kinder essen pizza" in Schweizerdeutscher Gebärdensprache?". Below the search bar, there are navigation links for "All", "Videos", "Images", "Books", "News", and "More". The search results show "About 623,000,000 results (0.46 seconds)". The main content area is split into two panels. The left panel is labeled "GERMAN" and contains the text "kleine kinder essen pizza". The right panel is labeled "SWITZERLAND" and contains a red line-art illustration of a person's face and upper body, representing a sign language gesture. At the bottom of the left panel, there are icons for a microphone and a speaker, and the text "25 / 500".

# sign.mt: Effortless Real-Time Sign Language Translation

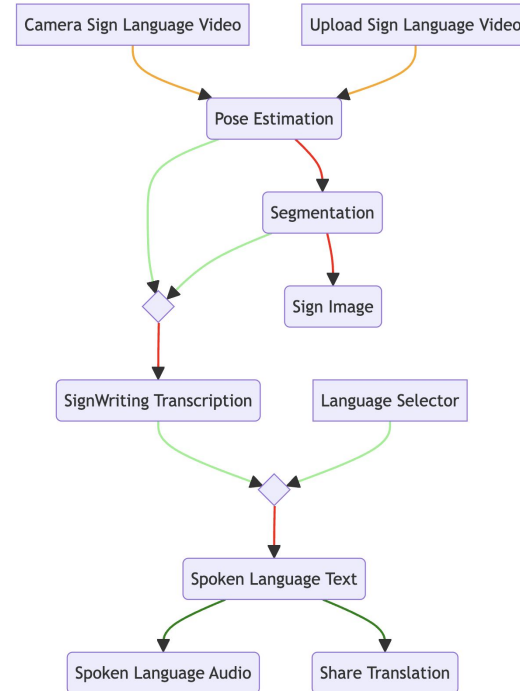
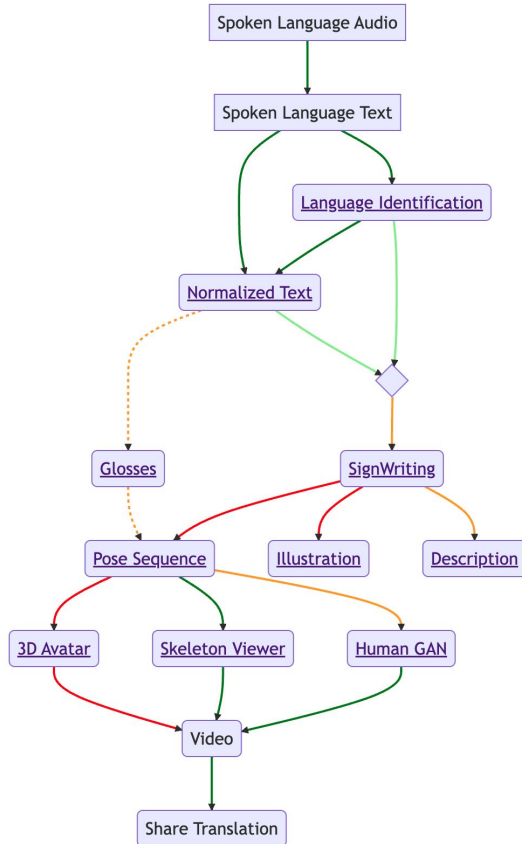


COUNTRY	USERS
United States	6.8K
India	1.7K
United Kingdom	963
Switzerland	893
Philippines	593
Germany	581
Canada	507



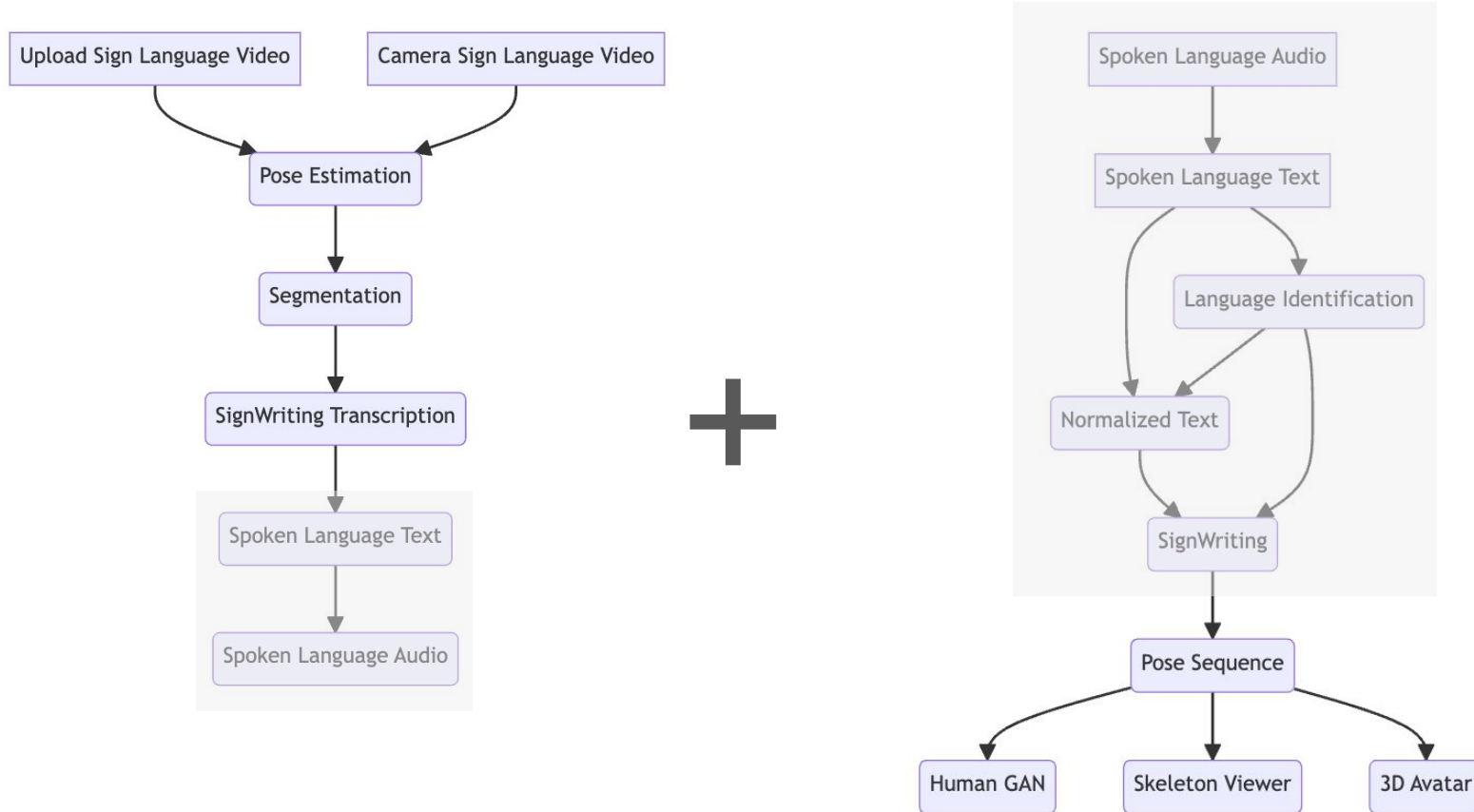
- Moryossef, A. (2023). [sign.mt](https://github.com/sign/translate): Real-Time Multilingual Sign Language Translation Application. <https://github.com/sign/translate>

# sign.mt: Implemented Pipelines





# Bonus Pipeline: Sign Language Anonymization



## Sign Language Processing

### Introduction

(Brief) History of Signed  
Languages and Deaf Culture

Sign Language Linguistics  
Overview

Sign Language Representations

### Tasks

Sign Language Detection

Sign Language Identification

Sign Language Segmentation

Sign Language Recognition,  
Translation, and Production

Fingerspelling

Annotation Tools

### Resources

Collect Real-World Data

Practice Deaf Collaboration

Downloading

Other Resources

# *pose-format*: Library for viewing, augmenting, and handling .pose files



```
from pose_format import Pose

data_buffer = open("file.pose", "rb").read()
pose = Pose.read(data_buffer)

numpy_data = pose.body.data
confidence_measure = pose.body.confidence
```

- Moryossef, A., Müller, M., & Fahrni, R. (2023). pose-format: Library for viewing, augmenting, and handling .pose files. <https://github.com/sign-language-processing/pose>

# sign-language-datasets

```
import tensorflow_datasets as tfds
import sign_language_datasets.datasets

# Loading a dataset with default configuration
aslg_pc12 = tfds.load("aslg_pc12")

# Loading a dataset with custom configuration
from sign_language_datasets.datasets.config import SignDatasetConfig
config = SignDatasetConfig(name="videos_and_poses256x256:12",
                           version="3.0.0",           # Specific version
                           include_video=True,        # Download, and load dataset videos
                           fps=12,                   # Load videos at constant, 12 fps
                           resolution=(256, 256),     # Convert videos to a constant resolution, 256x256
                           include_pose="holistic")   # Download and load Holistic pose estimation
rwth_phoenix2014_t = tfds.load(name='rwth_phoenix2014_t', builder_kwargs=dict(config=config))
```

- Moryossef, A., & Müller, M. (2021). Sign language datasets. <https://github.com/sign-language-processing/datasets>

# signwriting: Python utilities for SignWriting.

## Utilities

### signwriting.formats

This module provides utilities for converting between different formats of SignWriting. We include a few examples:

1. To parse an FSW string into a [Sign](#) object, representing the sign as a dictionary:

```
from signwriting.formats.fsw_to_sign import fsw_to_sign

fsw_to_sign("M123x456S1f720487x492")
# {'box': {'symbol': 'M', 'position': (123, 456)}, 'symbols': [{'symbol': 'S1f720', 'position':
```

2. To convert a SignWriting string in SWU format to FSW format:

```
from signwriting.formats.swu_to_fsw import swu2fsw

swu2fsw('M123x456S1f720487x492')
# M525x535S2e748483x510S10011501x466S2e704510x500S10019476x475
```

### signwriting.tokenizer

This module provides utilities for tokenizing SignWriting strings for use in NLP tasks<sup>[1]</sup>. We include a few usage non-exhaustive examples:










1. To tokenize a SignWriting string into a list of tokens:

```
from signwriting.tokenizer import SignWritingTokenizer










tokenizer = SignWritingTokenizer()

fsw = 'M123x456S1f720487x492S1f720487x492'
tokens = list(tokenizer.text_to_tokens(fsw, box_position=True))
# ['M', 'p123', 'p456', 'S1f7', 'c2', 'r0', 'p487', 'p492', 'S1f7', 'c2', 'r0', 'p487', 'p492']
```

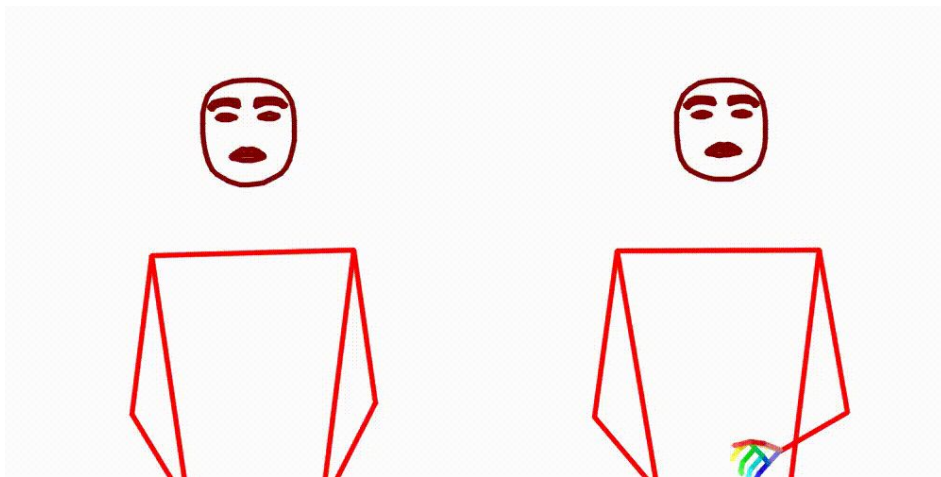
# signwriting-illustration: Automatic Illustration of signs written in SignWriting

	00004	00007	00015
Video			
SignWriting			
Illustration			
Prompt	An illustration of a person with short hair, with black arrows.	An illustration of a woman with short hair, with black arrows.	An illustration of a man with short hair. The arrows are black.

# signwriting-description: Describe how to perform signs from SignWriting

SignWriting	Translation	Description
	Hello	With your dominant hand open, touch your forehead and move your hand away, palm facing out.
	Thank You	Touch your dominant open hand to your lips, then move your hand forward, palm up.
	Help (him/her)	Place your dominant hand's fist (thumb up) on the palm of your open non-dominant hand. Move both hands upward together.
	No	With your dominant hand, extend your index and middle fingers while keeping your other fingers tucked in. Tap these fingers against your thumb.
	No	Shake your head horizontally while furrowing your eyebrows.
	Sorry	Form a fist with your dominant hand, palm facing in. Circle it over your heart.
	Friend	Link the index fingers of both hands together, alternating their positions.
	Love	Cross your arms over your chest as if giving yourself a hug, with your hands forming fists.
	Name	With your dominant hand, extend your index and middle fingers. Tap these fingers twice onto the extended index finger of your non-dominant hand, which is held horizontally.

# *sign-vq*: Vector Quantizer for Sign Language MediaPipe Poses






















# *signbank-plus*: Sign language translation dataset using SignWriting

Dataset	Training Pairs	Vocab	Sockeye		Fairseq		OpenNMT		Keras (mT5)	
			BLEU	chrF	BLEU	chrF	BLEU	chrF	BLEU	chrF
Original	521,390	6,016	0.2	8.4	0.18	4.74	0.69	9.21	0.07	6.39
Cleaned	357,574	5,200	<b>22.32</b>	<b>28.63</b>	1.1	<b>7.59</b>	<b>30.6</b>	<b>22.46</b>	<b>6.02</b>	12.35
Expanded	1,027,418	5,976	0.55	7.22	<b>1.26</b>	6.52	13.38	13.0	2.99	<b>12.49</b>

Table 1: Evaluation of the usability of our data for machine translation.

# signwriting-evaluation: Automatic Evaluation for SignWriting Machine Learning Outputs

			
CLIPScore	SymbolsDistances	TokenizedBLEU	CHRF
			
			
			
			

# Sign Language Alignment

00:08.34 Wenn euch jemand berührt  
00:12.42 oder aufdringlich wird, sagt zuerst freundlich Nein.

00:13.26 Aber wenn die Person dann weitermacht,  
00:17.11 muss man ihr klar und deutlich Nein sagen.

00:20.53 (Laut) Non!  
00:22.07

00:22.56 Ein scharfes Nein,  
00:25.60 kein langgezogenes Neiiiin - das bringt nichts.

00:25.60 Ein richtig scharfes Nein.  
00:27.60

00:30.33 Könnt ihr das?  
00:32.37

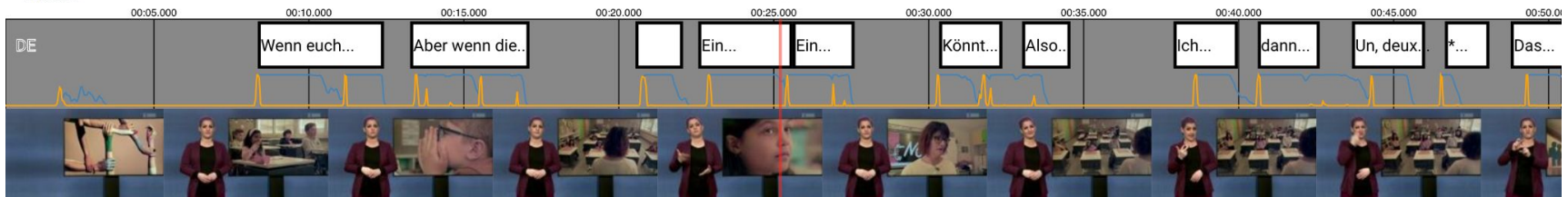
00:33.01 Also, steht mal bitte auf.  
00:34.57

00:37.89 Ich werde bis drei zählen,  
00:39.97

00:40.61 dann ruft ihr richtig hart: "Nein!"  
00:42.61



Ein scharfes Nein,  
kein langgezogenes Neiiiin - das bringt nichts.



that's all :)