Multimodal Sign Language Translator

FACULTY OF INFORMATION ENGINEERING, INFORMATICS AND STATISTICS

COMPUTER SCIENCE DEPARTMENT

MULTIMODAL INTERACTION

PROFESSOR MARIA DE MARSICO

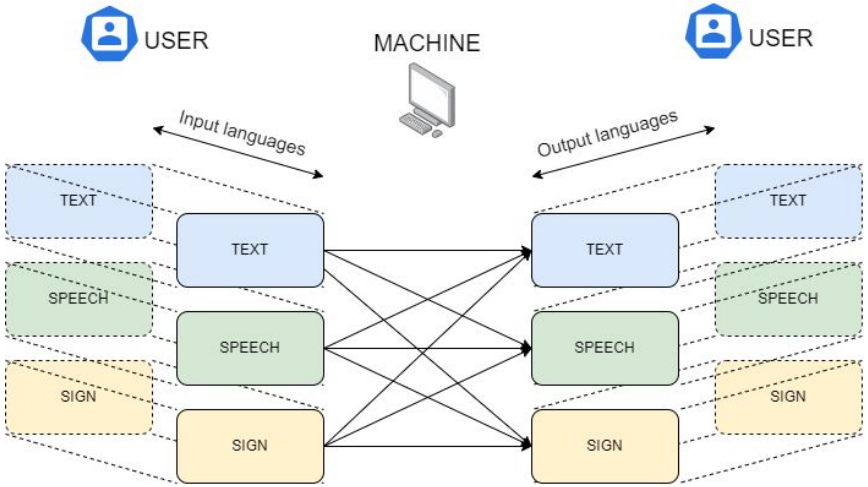
FEDERICO BARRECA

GIUSEPPE BELLO

ILARIA DE SIO

MATTEO MORTELLA

DANIELE SOLOMBRINO



High-level view

Sign language (SL) translations

- **Multimodal**
  - Tactile, Speech and Visual
- **Polyglot**
  - 🇮🇹, 🇬🇧, 🇺🇸 and 🇪🇸
  - Spoken & Signed languages
- **Bidirectional**
  - {speech, text, sign} ⟷ {speech, text, sign}

MMST development timeline

- **Domain** focus: sign language
- Domain **understanding** and knowledge gathering
- Real-world scenarios and **use-cases**
- **Theoretical** modeling and design
- **Practical** modeling and design

Real-world scenarios

Conversations between

- Two people using two different Sign Languages
- Person unable to use Sign Language, person only able to use Sign Language

No need for a physical **interpreter**

Conversion of material to sign language

Sign Language **learning**

- Test current knowledge
- Learn new signs

From scenarios to actual use-cases

Text-to-Sign translation

Speech-to-Sign translation

Sign-to-Text translation

Text-to-Speech translation

Text-to-Text translation

Sign-to-Sign translation

Speech-to-Speech translation

All modalities are valid and applicable in reverse as well!

Actors

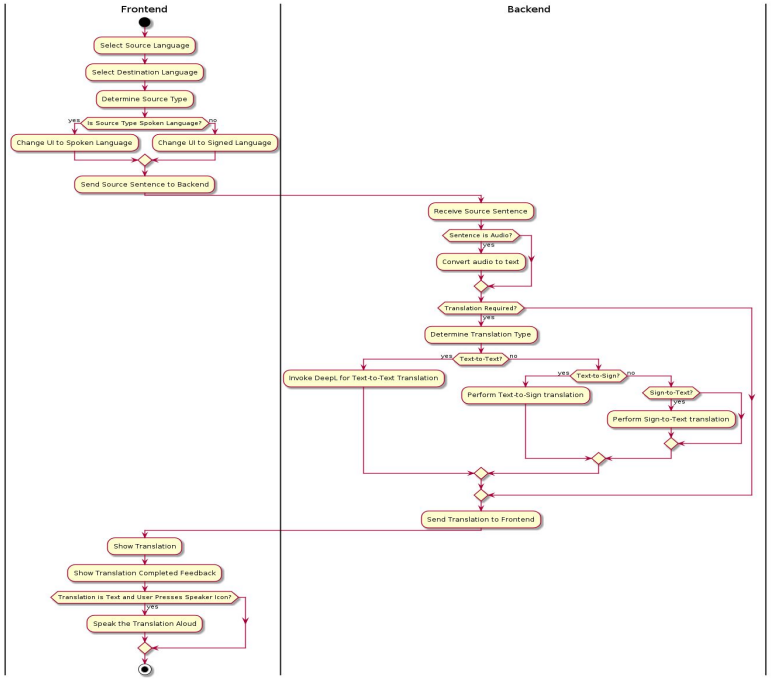
Preconditions

Main flow

Postcondition

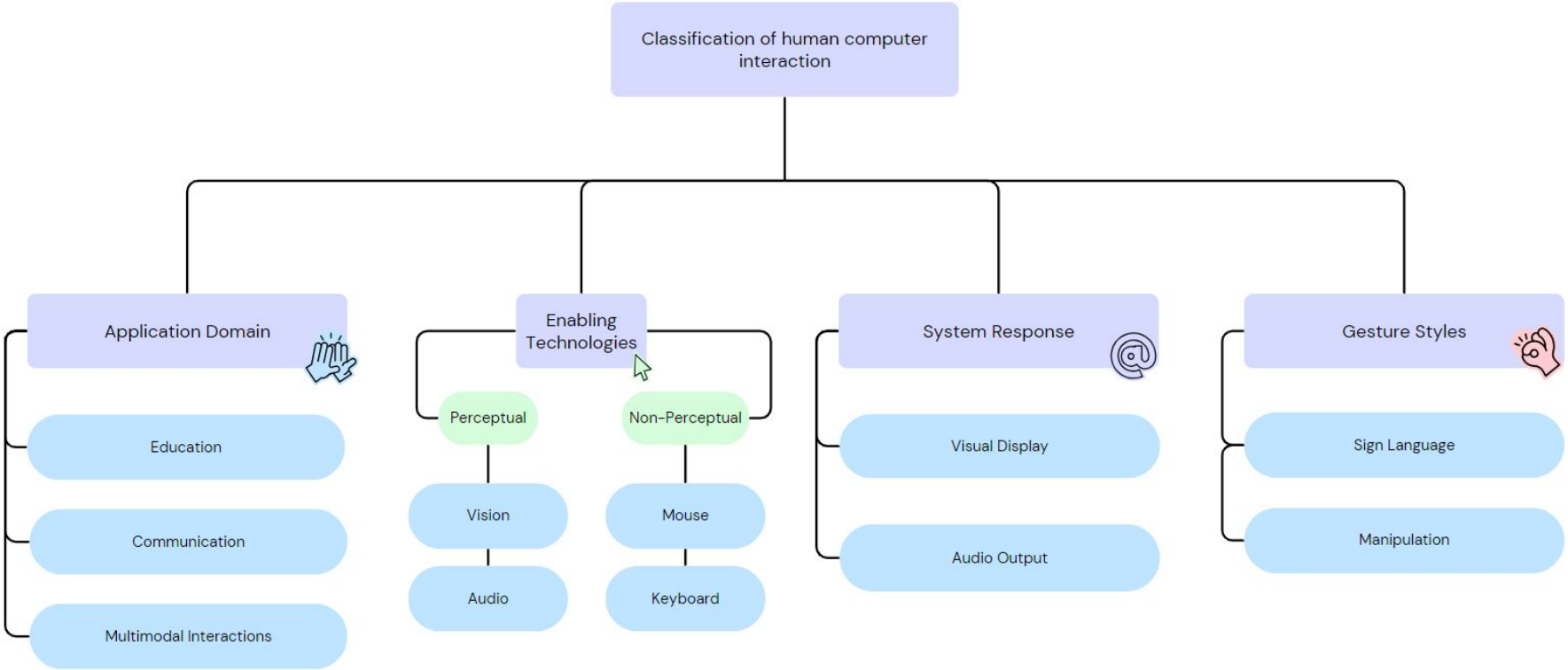
Theoretical Design Modeling

Activity diagram




```
@startuml
|Frontend|
start
:Select Source Language;
:Select Destination Language;
:Determine Source Type;

if (Is Source Type Spoken Language?) then (yes) :Change UI to Spoken Language;
else (no) :Change UI to Signed Language;
endif

:Send Source Sentence to Backend;

|Backend|
:Receive Source Sentence;

if (Sentence is Audio?) then (yes) :Convert audio to text;
endif

if (Translation Required?) then (yes) :Determine Translation Type;
  if (Text-to-Text?) then (yes) :Invoke DeepL for Text-to-Text Translation;
  else (no)
    if (Text-to-Sign?) then (yes) :Perform Text-to-Sign translation;
    else (no)
      if (Sign-to-Text?) then (yes) :Perform Sign-to-Text translation;
      endif
    endif
  endif
endif

:Send Translation to Frontend;

|Frontend|
:Show Translation;
:Show Translation Completed Feedback;

if (Translation is Text and User Presses Speaker Icon?) then (yes)
  :Speak the Translation Aloud;
endif
stop
@enduml
```

Human-Computer Gesture Interaction



Human-Computer Interaction

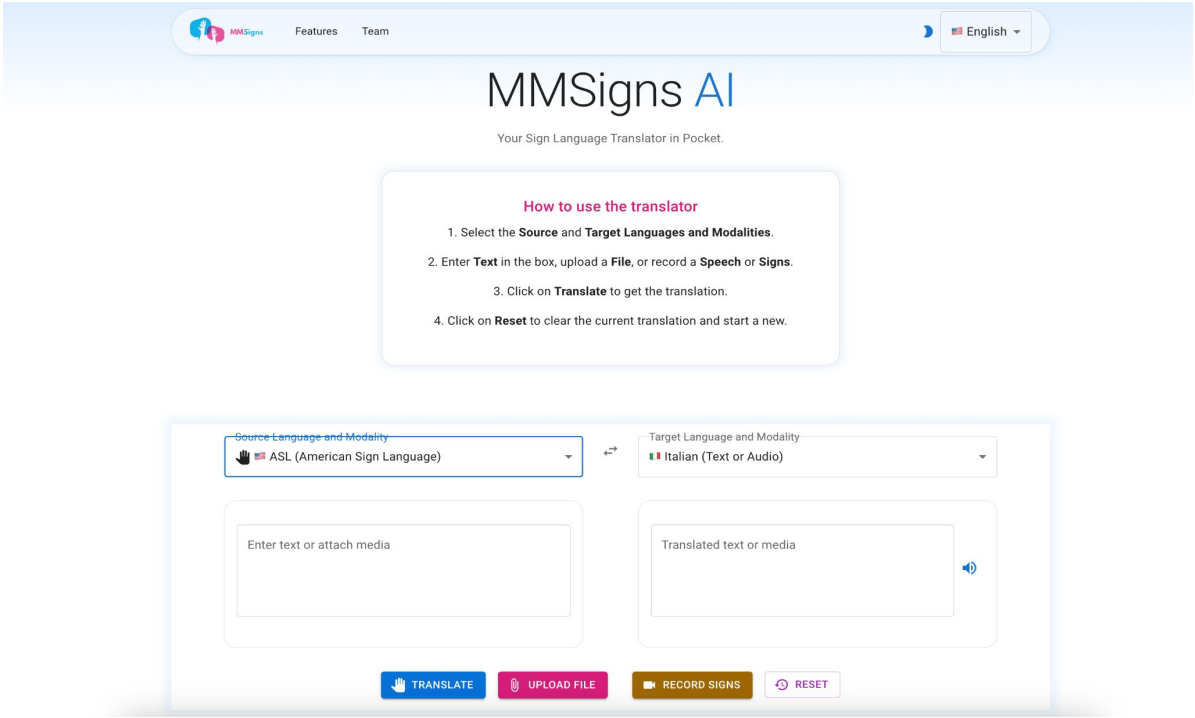
User Interface

- **Usable**
  - Requires < 2 mins to learn
- **Accessible**
  - Multi-language support
  - Comply with accessibility requirements
    - **WCAG 2.2**., A and AA (certain cases)

User Experience

- **Intuitive**
  - Buttons, drop down menus, text boxes
- Step-by-step **instructions**
- Browser- and OS-**agnostic**

Human-Computer Interaction



SAPIENZA
UNIVERSITÀ DI ROMA

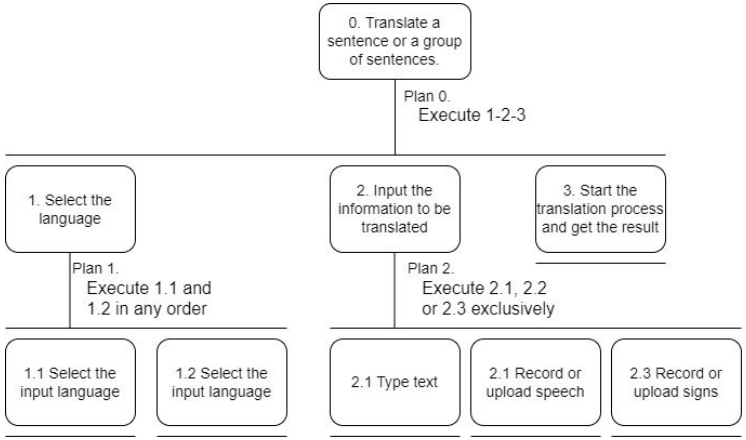
Separable Interface Design

Theoretical **desideratum** → keep presentation and functionality separate

Practical implementation → stay tuned…

Task Analysis

- **H**ierarchical **T**ask **A**nalysis
  - CTT logical-temporal operators not required
- Three-level **hierarchy**
  - Task → Sub-tasks → Sub-sub-tasks
- Three **plans**
  - Fixed
  - Discretionary
  - “Alternative”



Multimodal Coordination

- **Alternative/Equivalent** (Bernsen and Dybkjær /Martin) → inputs
  - Semantic content **invariant** w.r.t. modality
    - “apple”, [pronunciation of “apple”], [sign language video of “apple”] → always refers to 🍏
- **Transfer**
  - In some cases, input modality ≠ output modality

- **No Specialization**
  - Incompatible with full equivalency

SAPIENZA
UNIVERSITÀ DI ROMA

Multimodal Fission

- **Fission** (Foster) ⟶ outputs
  - **Implicit** message ⟶ translation done
  - Distributed across multiple channels ⟶ audio and visual **feedbacks**
    - Multimodal feedbacks used to **highlight new content** and **explain context**

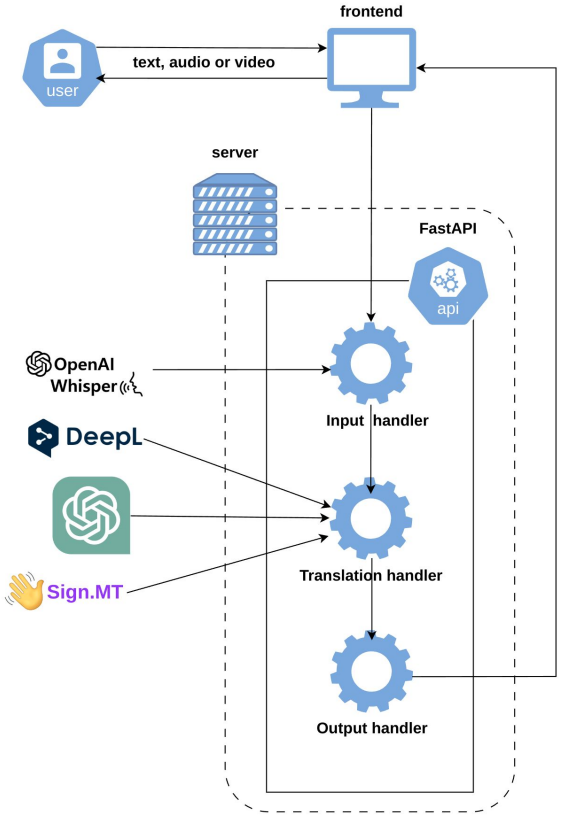
  - **Gestalt Laws**
    - **Proximity, Closure and Continuity** laws to avoid artifacts and ambiguity
    - Used **Similarity** law to separate different components, for example closed box to distinguish “instructions block” and “translation block”

Practical Design Modeling

Separable Interface Design

Theoretical desiderata → keep presentation and functionality separate

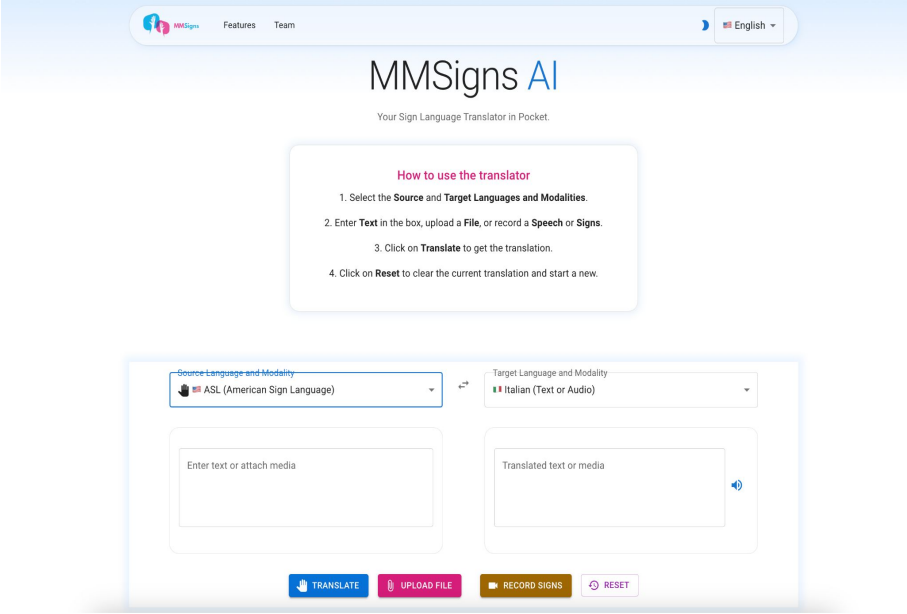
Practical **implementation** → now! 🎉



Separable Interface Design

Seehein Model

- Presentation component
  - **Frontend**
- Application interfaces components
  - **Backend**
- Dialogue component
  - Frontend⟵ ⟶backend HTTPS requests



Frontend

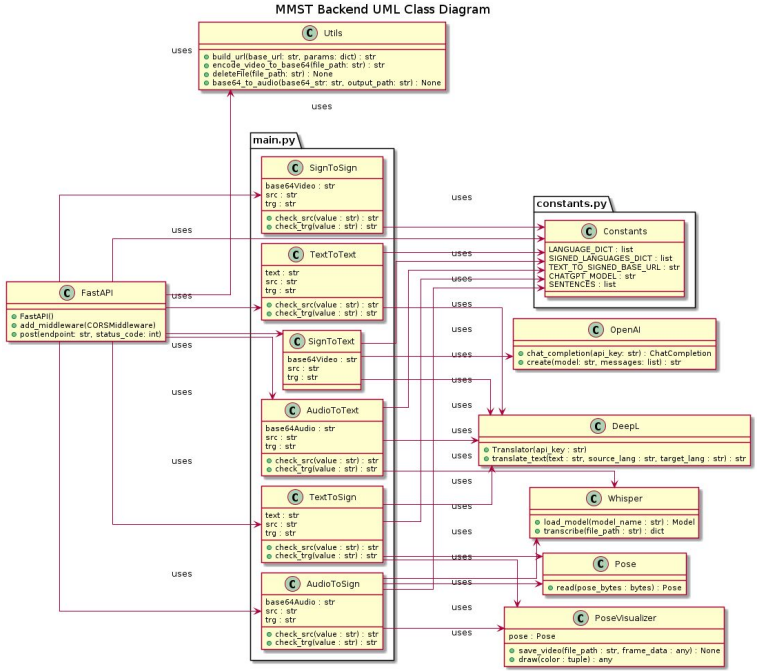
Presentation component

- Show app **logo** and **tutorial**
- Select source and destination languages
- **Input** source sentence
- **Output** translation
- Select preferred **modality**

Backend

Application component

- Input **pre-processing**
- **Calls** to external services
  - OpenAI Whisper, ChatGPT, DeepL, Sign.mt
- Output **post-processing**



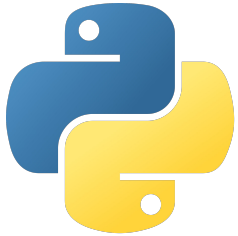
Frontend ⟷ Backend

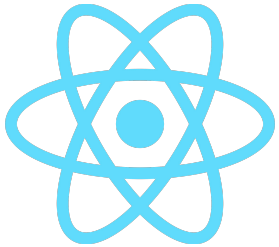
Dialog component

- **Bidirectional** communication
  - Frontend ⟶ Backend
  - Backend ⟶ Frontend
- Input **validation**
- **API** entry points
  - {audio, text, sign}-to-{audio, text, sign}

```python
@app.post("/translate/text_to_text", status_code=200)
async def text_to_text(req: TextToText):
    translator = deepl.Translator(os.getenv("DEEPL_API_KEY"))
    text_info = translator.translate_text(
        req.text,
        source_lang=req.src,
        target_lang="en-us" if req.trg == "en" else req.trg,
    )
    return {"result": str(text_info)}
```

Frameworks and libraries





Future Works

- Expand
  - Multimodal **coordination**
    - Complementarity & redundancy
    - Ambiguous translations & reinforce SL concepts
  - More signed and spoken languages
  - Mobile and tablet apps
- Include
  - **Real-time** capabilities → AR overlay of SL translations
  - Natural Language **Understanding**

Conclusion and Recap

- **Sign language** translations
- **Multimodal**
  - Tactile, Speech and Visual
- **Polyglot**
  - 🇮🇹, 🇬🇧, 🇺🇸 and 🇪🇸
  - Spoken & signed languages
- **Bidirectional**
  - {speech, text, sign} ⟷ {speech, text, sign}

- Activity Diagram and use-cases
- UI and UX design
- Separable Interface Design
  - Backend, Frontend and Communication
- Task Analysis
  - Task Hierarchy
  - Fixed, discriminative, “alternative” plans
- Multimodal coordination
  - Alternative/Equivalent and Fission
  - Complementarity and Redundancy 🔮

SAPIENZA
UNIVERSITÀ DI ROMA

THANKS FOR THE ATTENTION!

MMST – Multi**m**odal **S**ign **L**anguage **T**ranslator