

Towards AI-driven Sign Language Generation with Non-manual Markers

Han Zhang*
University of Washington, USA
micohan@cs.washington.edu

Connor Gillis
Apple, USA
connorgillis@apple.com

Lorna Quandt*
Gallaudet University, USA
lorna.quandt@gallaudet.edu

Rotem Shalev-Arkushin*
Tel-Aviv University, Israel
rotems7@mail.tau.ac.il

Gierad Laput
Apple, USA
gierad@apple.com

Leah Findlater
Apple, USA
lfindlater@apple.com

Colin Lea
Apple, USA
colin_lea@apple.com

Vasileios Baltatzis
Apple, USA
vbaltatzis@apple.com

Raja Kushalnagar*
Gallaudet University, USA
raja.kushalnagar@gallaudet.edu

Abdelkareem Bedri
Apple, USA
bedri@apple.com

Outline

- Introduction
- Method
- Evaluation
- Discussion

Introduction

อุปสรรคในการสื่อสารระหว่างคนหูหนวกกับคนปกติทั่วไป

- ขาดล่ามภาษามือที่มีทักษะอย่างจำกัด
- บุคคลทั่วไปมีความสามารถในการใช้ภาษามือต่ำ
- เทคโนโลยีต่างส่วนใหญ่เป็นภาษาพูดหรือภาษาเขียน



Figure 1: รายการข่าวพร้อมล่ามแปลภาษามือ

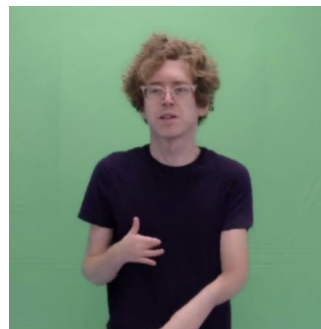


Figure 2: การสังเคราะห์ล่ามแปลภาษามือ 1

Introduction

ปัญหาหลักในการสังเคราะห์ภาษามือ

- อวาตาร์ที่มีคุณภาพต่ำ
- การแปลที่ไม่แม่นยำ
- การลดความซับซ้อนหลักการภาษามือเช่นสีหน้าและท่าทางประกอบการสื่อสาร



Figure 3: การสังเคราะห์ล้ามแปลภาษามือ 2

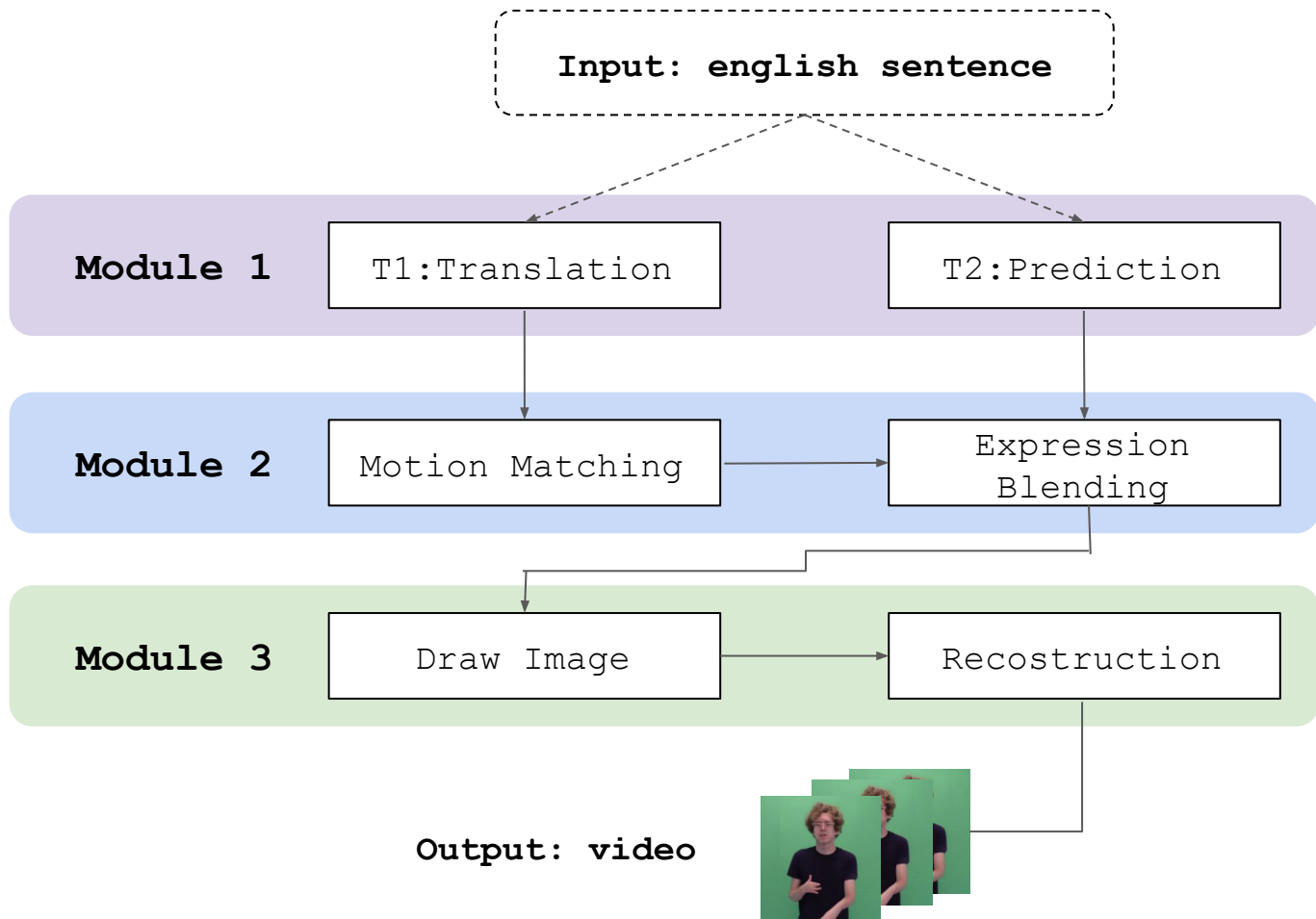
Introduction

ทำไมสีหน้าและท่าทางจึงสำคัญในภาษามือ

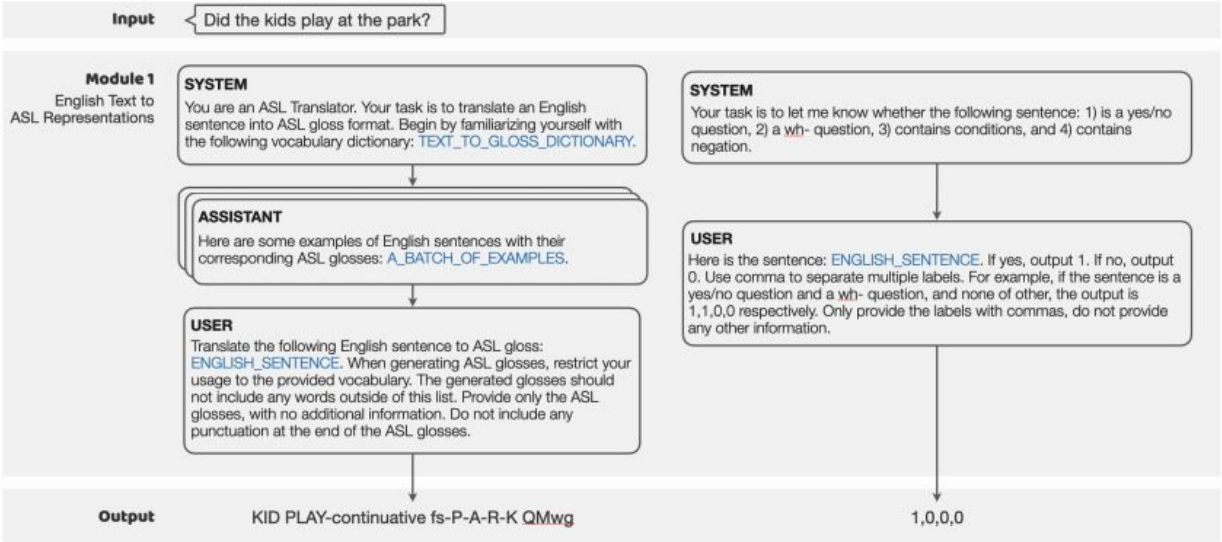


Figure 4: การเปรียบเทียบสีหน้าในแปลภาษามือคำเดียวกัน

Method



Method (Module 1)



Model
GPT-2
GPT-3.5-turbo-0125
GPT-4-turbo-2024-04-09
GPT-4-0125-preview
GPT-4o-2024-05-13 (Our adopted model)

- 1. Yes-no question?
- 2. Wh-question?
- 3. Conditional statement?
- 4. มี negation ไหม?

Figure 5: แสดงการทำงานของโมดูล 1

Method (Module 2)

คำถามแบบ yes/no → ยกคิ้ว

คำถามแบบ wh-question → ขมวดคิ้ว

หรือไม่ใช่คำถาม → คิ้วปกติ

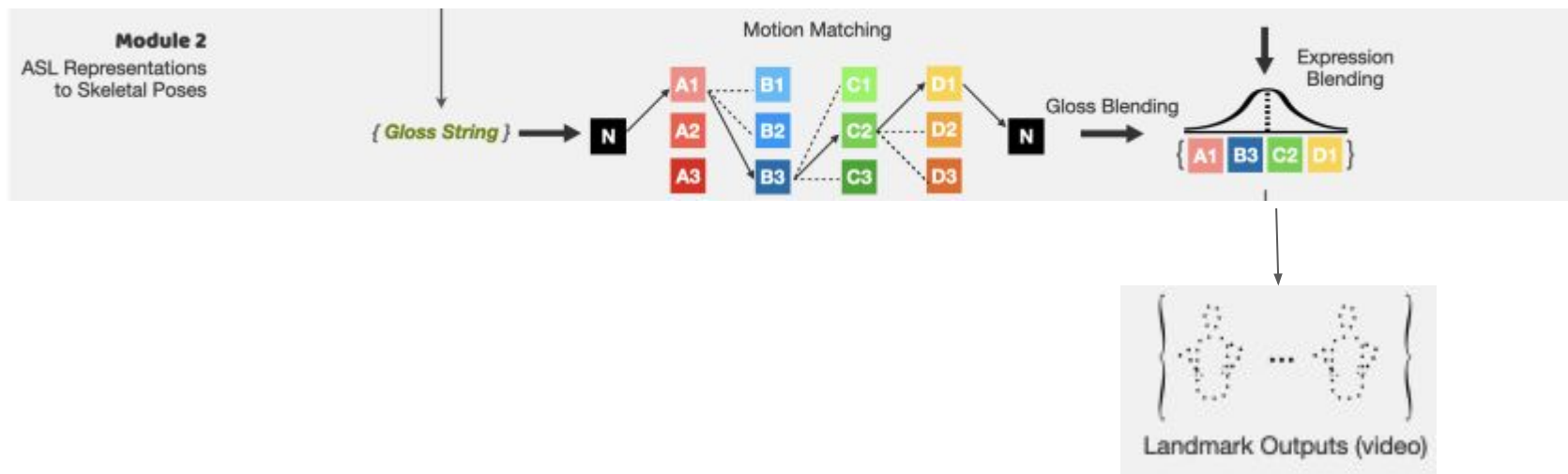


Figure 6: แสดงการทำงานของโมดูล 2

Method (Module 3)

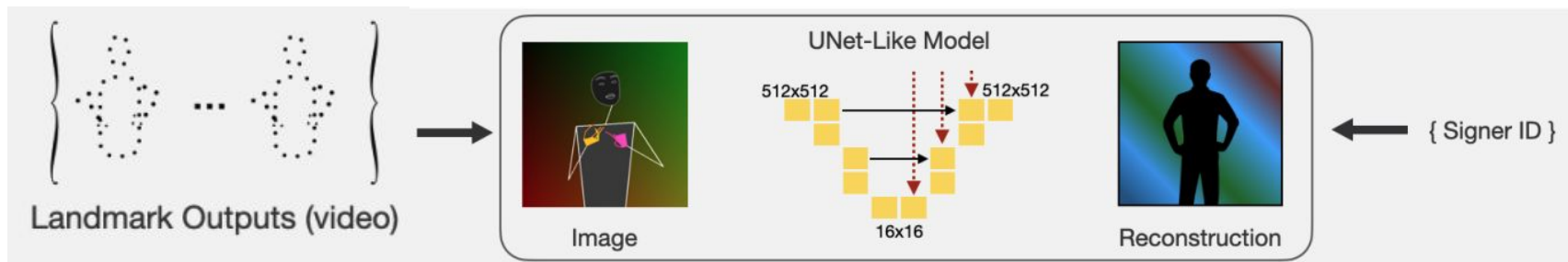
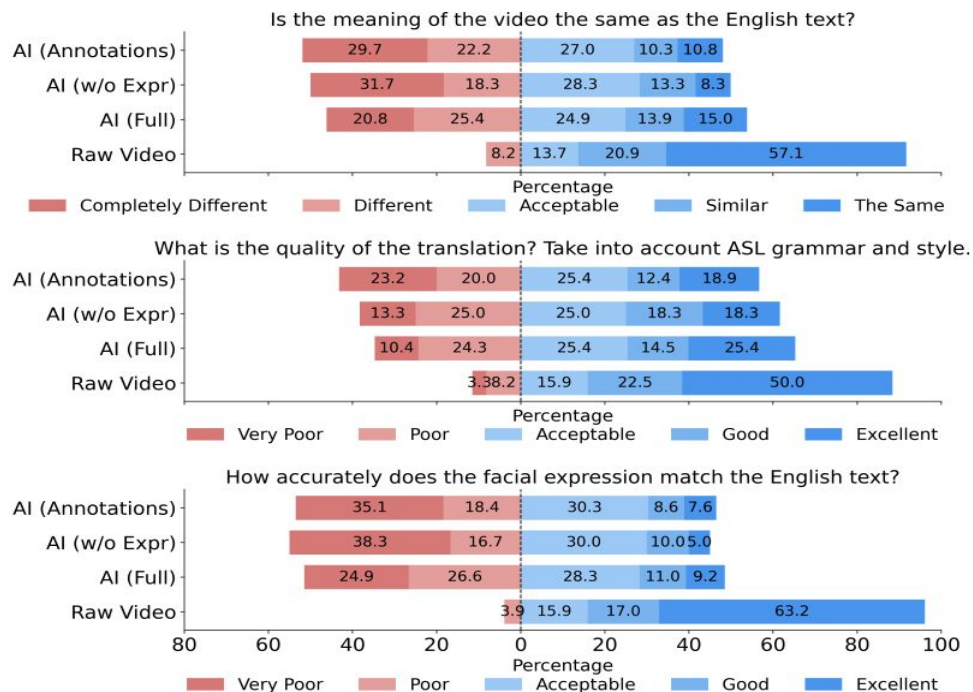


Figure 7: แสดงการทำงานของโมดูล 3

Evaluation

(Annotation) = no-llm input for m1,m2

(w/o Expr)= expression blending model turned off



(b) Translation Quality.

Discussion

Goal 1: พัฒนาด้านแบบระบบการสร้างภาษามืออเมริกัน (ASL) เพื่อแก้ไขข้อจำกัดที่ยังคงเป็นอุปสรรคในการสร้างระบบแปลภาษามืออัตโนมัติ

Goal 2: เพื่อศึกษาว่าคนหูหนวกมีมุมมองอย่างไรต่อผลลัพธ์ที่ระบบสร้างขึ้น

Results: มากกว่าครึ่งของผู้เข้าร่วมทดสอบพบว่าความหมายของวิดีโอที่สร้างขึ้นนั้น**ยอมรับได้** เมื่อเทียบกับข้อความต้นฉบับ อย่างไรก็ตามในตัวอย่างทดสอบเกือบ 50% ผู้ทดสอบให้คะแนนการแปล **แย่ หรือแย่มาก** ในด้านไวยากรณ์และรูปแบบภาษามือ

จุดสำคัญของงานวิจัยนี้คือ **การใช้ zero-shot prompting เพื่อดึงสีหน้าและท่าทางจากข้อความอังกฤษ ที่ช่วยให้วิดีโอภาษามือดูเป็นธรรมชาติมากขึ้นและถูกต้องตามหลักไวยากรณ์** อย่างไรก็ตาม ยังพบว่าลักษณะทางภาษาบางประการถูกตีความหมายผิด ซึ่งเกิดจากความไม่สอดคล้องกันระหว่างผลลัพธ์ข้อความที่ได้จาก LLM ในการแปลภาษาและข้อความต้นฉบับ อันเป็นอีกหนึ่งความท้าทายที่ต้องพัฒนาให้ดีขึ้นในการออกแบบระบบแปลภาษามือในอนาคต

References

1. <https://arxiv.org/pdf/2502.05661>
2. <https://www.youtube.com/watch?v=FTNxR72CXTg>
3. <https://signavatars.github.io/>