DragonFly

ASL- English MT

October 9, 2020

2020 AMTA Virtual Conference
Government Track

DragonFly

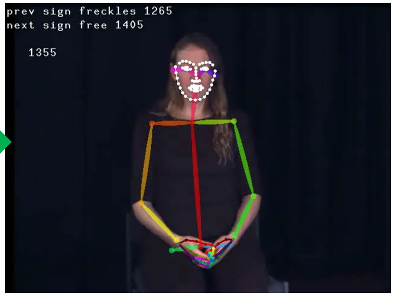
- Technology to enable the deaf and hearing to seamlessly communicate with one another without the assistance of an interpreter
  - Automated Machine Translation (MT) capabilities for enabling communication between speakers of American Sign Language (ASL) and English
  - Dragonfly will operate on the majority of IOS and ANDROID wearable devices including smart phones, tablets, and smart watches

Face-to-Face… Naturally… Anytime… Anywhere…

ASL Signer to English

Smart device captures the video of the sign

Sign is processed and recognized by DragonFly



ASL is translated into English

FREE

Text is displayed and the audio is voiced aloud

FREE

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 - 9, 2020, Volume 2: MT User Track

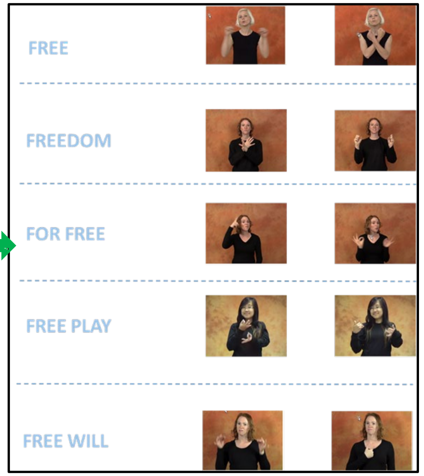
English Speaker to ASL

Spoken voice is captured by DragonFly

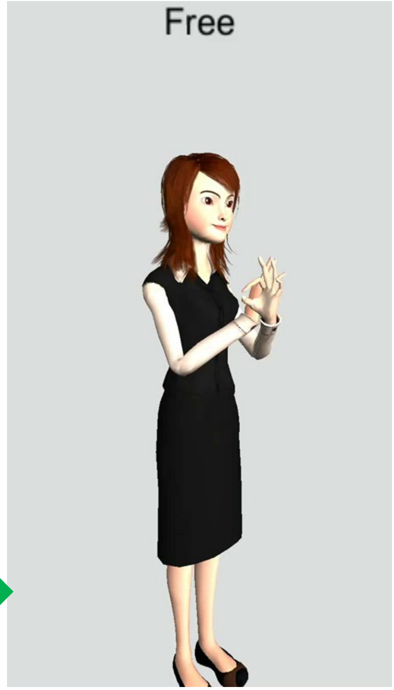
Speech is converted to English text

FREE

English to ASL translation performed



Text and Signing Avatar displayed to user



ASL Translation Challenges

Sign/Signer Variability

Distinct language with broad sign variation across signers

Signal Complexity

- Head/Face Gesture
- Handshape
- Hand Orientation
- Hand Placement
- Hand Motion

Sensor Variability

e.g. 2D/3D, fixed/mobile sensors

Session Variability

e.g. observation angle

Data Availability

- Limited availability of well annotated ASL<->English content

5

What we did

- Integrated the ASL Recognizer (ASLR) into an automated MT system that can be used in real time ad-hoc communications between signers and non-signers

- Implemented deep learning-based models (in OpenMT)
  - ASL video-to-ASL symbol sequence classifier
  - ASL symbol sequence-to-English sentence generator
  - ASL video-to-English sentence generator

What we did - Continued

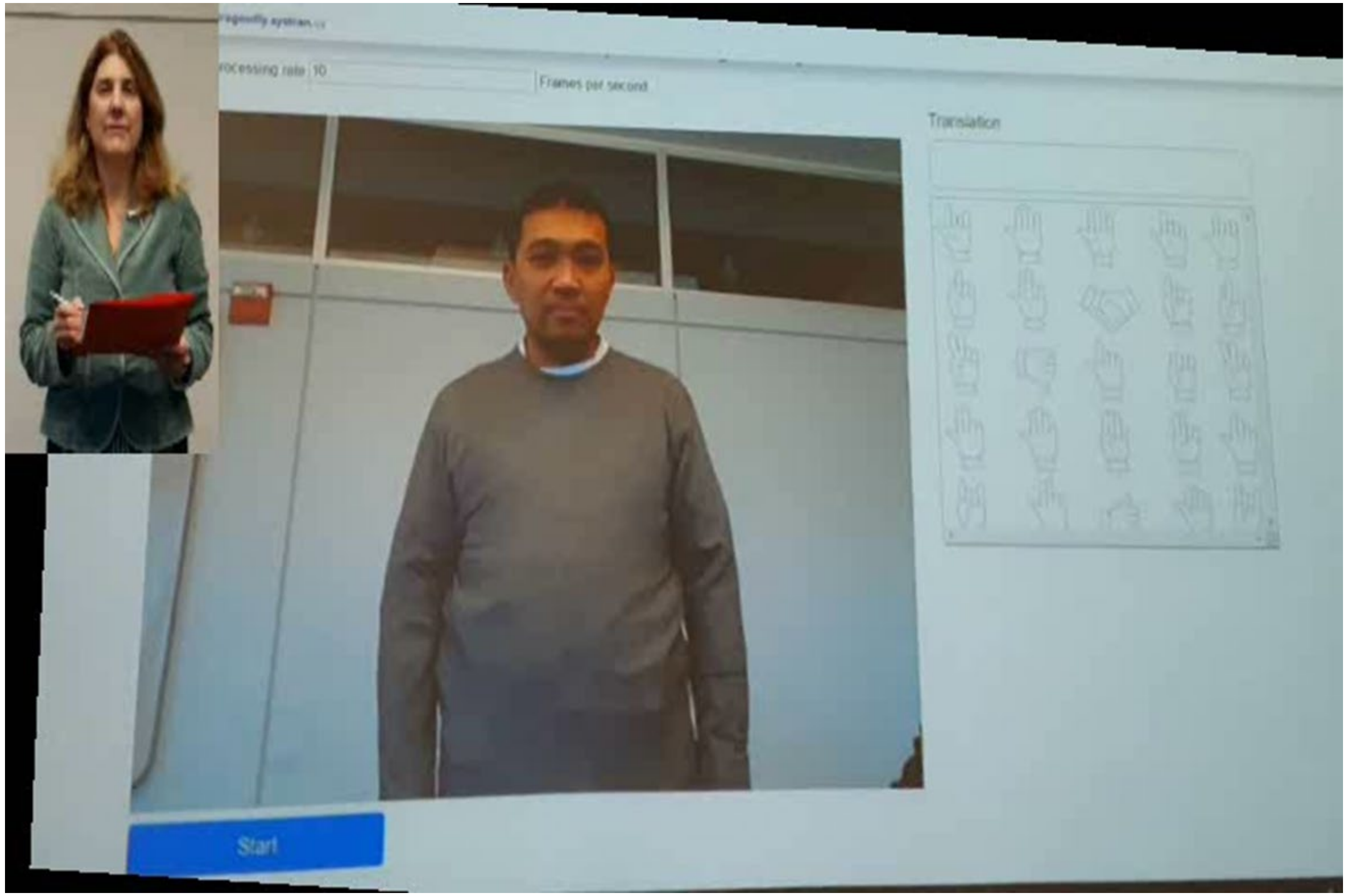
- Leveraged sources in addition to BU and Purdue data (e.g. closed-captioned ASL from The Sign Language Channel)
- Created and incorporated the use of computer-generated Synthetic Data to augment training data
- Development and testing of a Handheld Prototype

Handheld Prototype

- Web-based application
- ASL Video captured in real time on Smartphone or Tablet
- Seamlessly transmitted to Amazon Cloud for processing
- English MT text delivered and displayed in chat window
- Text-to-speech performed locally

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 - 9, 2020, Volume 2: MT User Track

Cafe DragonFly Demo



What we learned

- Computer generated Synthetic Data improved overall ASLR performance
  - Per sign raw recognition improved over 10% in initial testing
  - Key driver is using “valid” synthetic data to train the models
- Dramatically improved the speed and accuracy of ASLR
- However, we encountered both classic neural net and synthetic data validation challenges

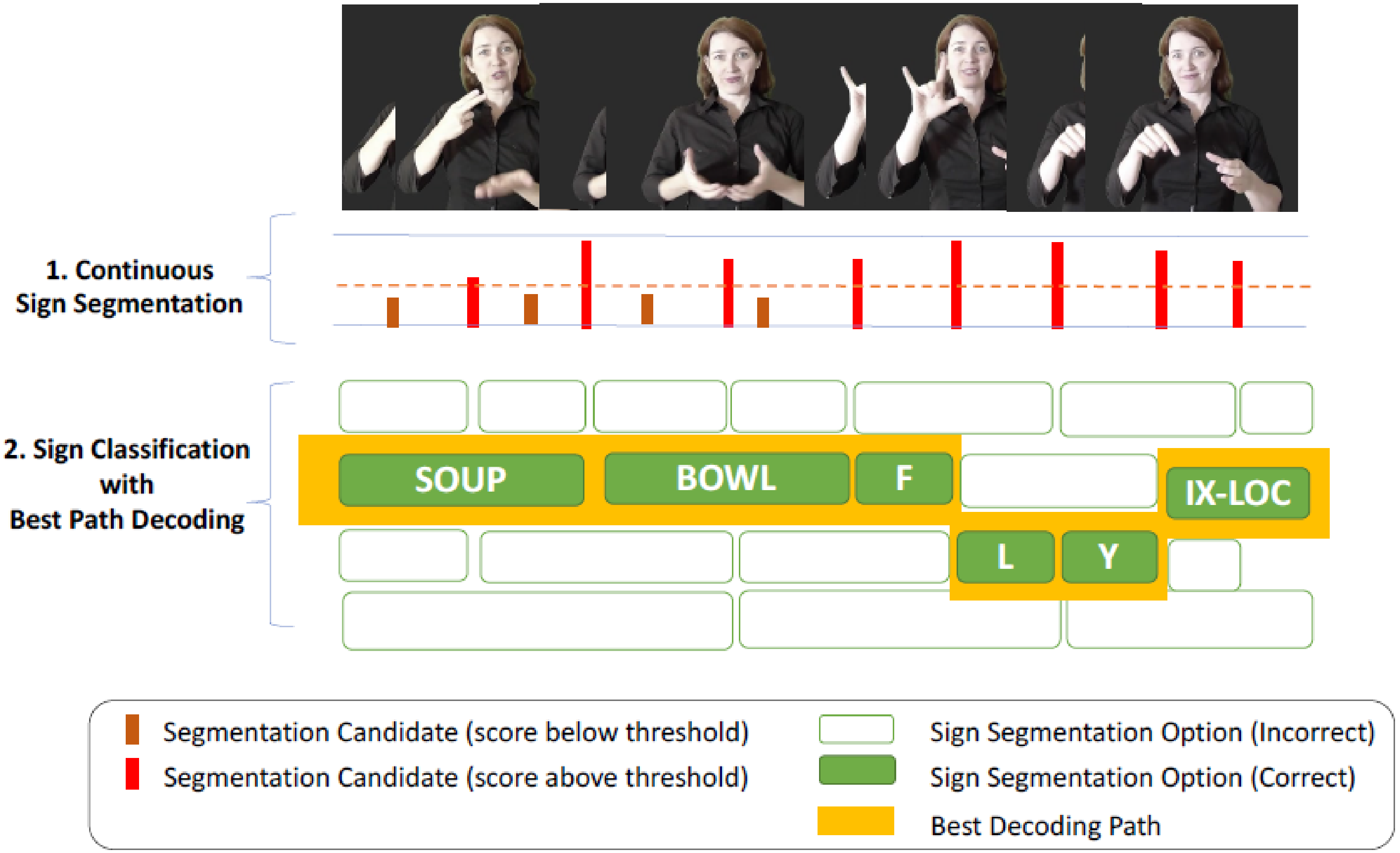
Neural Net Challenges

- MT Output Errors
  - Error Types typical of neural net encoder-decoder architecture model for low resource languages (Vardaro et. al 17 September 2019:Informatics, Koponen et.al 2019: Machine Translation).
    - Substitution
      - Input: “I want hamburger without mayonnaise, please.”
      - Output: “I will have a chicken sandwich without mayonnaise please.”
    - Reordering /Addition
      - Input: “I want cheeseburger and soup, please.”
      - Output: “I will have soup, cheeseburger and french fries please.”
    - Omission
      - Input: “I want hotdog with ketchup and mustard, please .”
      - Output: “I want hotdog with ketchup and mustard, __________.”
    - Addition /Substitution
      - Input: “I want cheeseburger with extra spinach .”
      - Output: “I want cheese pizza with extra spinach and bacon .”

Sequence-to-Sequence Challenges

- Lack of scalability: inability to easily add new words
  - Multiple training sentences required for each flavor of ice cream
    - I want chocolate ice cream
    - I want vanilla ice cream
    - I want strawberry ice cream, etc.
    - Some words characterized as only nouns so unable to say “peach Ice cream”
  - Multiple permutations required for each sentence instead of for individual words (want, would, love, like chocolate, vanilla, milk, pudding)
    - I want chocolate ice cream; I would like chocolate milk, I love chocolate,
    - I want chocolate pudding, I like vanilla pudding, I like chocolate milk, etc.

Continuous Sign Recognition



Continuous Sign Recognition Approach with Explicit Sign Segmentation and Sign Classification Steps

Sign Segmentation Process

(Khan 2014; Farag & Brock 2019), and

Future Plans

- Full scale development, test and evaluation of hand-held operational prototypes
- Platform (iOS, Android, and Windows) and browser (Chrome, Firefox, and Edge) compatibility and user field testing
- Incorporation of ASL avatar for signing synthesis
- Commercial partnerships for product delivery

Questions